Health and Incomes in Malawi and the KwaZulu-Natal Province of South Africa

Gemma Elizabeth Blackledge-Foughali

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Management and Languages

MAY 2016

Abstract

This thesis examines the relationship between health and income in Malawi and the KwaZulu-Natal province in South Africa.

The first empirical chapter considers the previously unused Second Integrated Household Survey (IHS2) to examine child mortality in Malawi. Household income and bed net use are included within the analysis presented in addition to the standard proximate determinants of population health. It is found that possessing at least some education and using bed nets significantly reduces child deaths.

The second empirical chapter explores the effect of morbidity upon daily wage rates in Malawi using the IHS2. The results suggest that even relatively short periods of morbidity reduce the daily wage received by relatively large amounts.

The final empirical chapter examines the impact of different types of “health shocks” upon household income in 1998 and 2004 using the KwaZulu-Natal Income Dynamics Study (KIDS) dataset. The results suggest that the effect of “health shocks” upon household income is negative in 1998 using a propensity score matching method. Direct income losses associated with shocks are also examined and indicate that for most households the effect of health shocks is generally small.

Dedication

In loving memory of my father-in-law Jamil Eddine Foughali. I will never forget you.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors, Professor Mark Schaffer and Professor Paul Hare. I am forever grateful for the opportunity I have had to benefit from their vast experience and knowledge.

I would like to offer my sincere appreciation for all the time my supervisors have given me. Such time has been invaluable and I have profited immeasurably from their advice and guidance. Most importantly, however, I would like to say thank you for their unwavering support, encouragement and patience throughout, which has made this thesis possible.

I would like to thank the ESRC, who funded the research undertaken within the thesis. I also would like to express my gratitude to Dr Andrew Clark and colleagues at the African Development Bank. The experience I had working with you is one I will never forget. I also wish to thank Antonio Nucifora at the World Bank, for his assistance in accessing the Malawi survey data used within this thesis.

I am also thankful for all the support given to me by my colleagues in the School of Management and Languages at Heriot-Watt University and the library staff. If I were to list everyone individually this acknowledgement would be many pages long. I therefore hope that all of you know who you are and how much you mean to me.

Almost last, but by no means least, I would like to say thank you to my friends. To my non-economist friends, thank you for listening and for the smiles and nods. To my economist friends, thank you for sharing your thoughts and suggestions. The Stata tips and tricks are all safety tucked away in notebooks for future use!

Finally, I would like to thank the very special people who I absolutely could not have done this without: my Mum and Tom, and my Dad and Jenni. I would also like to thank my colleague and very good friend Valerie Dickie: you have inspired me to be more than I ever thought I could be. Last but not least, to my amazing husband Callum: thank you for being my rock, for always believing in me when I didn’t, and standing by my side.

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	Gemma Elizabeth Blackledge-Foughali		
School/PGI:	School of Management and Languages		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought (Award **and** Subject area)	PhD Economics

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
E-thesis Submitted (***mandatory for final theses)***			
Signature:		Date:	

Table of Contents

List of Tables

List of Tables continued

List of Tables continued

List of Figures

Glossary

The following abbreviations are used throughout this thesis:

AIDS	Acquired Immunodeficiency Syndrome
ADL	Assessment of Daily Living
ART	Antiretroviral therapy
ATE	Average treatment effect
ATT	Average treatment effect on the treated
BMI	Body Mass Index
CIA	Conditional Independence Assumption
CME	Child Mortality Estimates
CSPRO	Census Survey and Processing System
DALYs	Disability-Adjusted Life Years
DHS	Demographic Health Survey
DID	Difference-in-differences
DSD	South African Department of Social Development
DTP3	Diphtheria-Tetanus-Pertussis
EA	Enumeration area
EPI	Extended Programme on Immunization
GTZ	German Agency for Technical Cooperation
HIV	Human Immunodeficiency Virus
IFPRI	International Food Policy Research Institute
IGME	Interagency Group for Child Mortality Estimation
IHS2	Second Integrated Household Survey
IV	Instrumental Variables
JCE	Malawi Junior Certificate Examination
KHDS	Kagera Health and Development Survey
KIDS	KwaZulu-Natal Income Dynamics Study
LAM	Lactational Amenorrhea Method
LSHTM	London School of Hygiene and Tropical Medicine
LSMS	Living Standards Measurement Survey
MDGs	Millennium Development Goals
MEC	Marginal Efficiency of Capital
MNT	Maternal and Neonatal Tetanus
MSCE	Malawi School Certificate of Education
NIBR	Norwegian Institute for Urban and Regional Studies
NN	Nearest Neighbour
OLS	Ordinary Least Squares
PSLC	Primary School Leaving Certificate
PSLSD	Project for Statistics on Living Standards and Development
PSM	Propensity Score Matching
QALYs	Quality Adjusted Life Years
SALDRU	South African Labour and Development Research Unit
TBA	Traditional Birthing Assistant
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNICEF	United Nations Children's Fund
WHO	World Health Organisation

Chapter 1: Introduction

Good health is something which many of us take for granted in developed countries. It is not until we experience morbidity that we consider the effect that it can have on our daily lives. We can turn on our taps and get clean water, take our children to the doctor to be vaccinated, and buy as much food as we possibly want in our supermarkets. But this is not how many people in the world live.

Comparisons of different measures of health across the globe show stark differences in the experiences of people living in high and low income countries. For example, estimates by the World Bank indicate that the life expectancy of a male living in Sub-Saharan Africa is just over 53 years, compared to just over 77 years in a high income country. For women, the outlook is fairly similar, with an expected lifespan of just over 55 years in Sub-Saharan Africa compared to almost 83 years in a high income country (World Bank, 2012). [1]

For people living in developing countries it is not a simple matter of opening a cupboard, picking up a telephone or driving down the street to see a doctor. For the world's most vulnerable people, good health can be hard to achieve and it starts with investments made during childhood.

Ensuring good health at the early stages of life is essential for the development of a strong and productive workforce. In order to achieve this, we need to have a clear understanding of the factors which must be in place to protect people from debilitating health conditions in the first instance. As medical science has progressed, the incidence of disease in many countries has fallen, resulting in greater life expectancy and more fulfilling lives.

There have also been considerable gains made in improving the livelihood outcomes that children face over the last 20 years. The significant reduction in the number of deaths occurring in children aged under five has primarily been attributed to

[1] 2010 figures in all cases.

improvements in access to clean drinking water, micronutrient supplements and rehydration therapy, along with large and widespread vaccination programmes.

Improving child health is at the heart of the Millennium Development Goals (MDG).[2] The MDG are a global commitment to improve the welfare and livelihoods of the world's poorest people by lifting them out of poverty by 2015.

A particular challenge within sub-Saharan Africa is the prevalence of HIV/AIDS. It is the worst affected region in the world: 69 percent of those living with HIV reside in sub-Saharan Africa (UNAIDS, 2012). It is estimated that an additional four million deaths have occurred within the region as a result of AIDS (United Nations, 2004). HIV/AIDS will continue to be a challenge to health for the foreseeable future, as it is believed that the number of deaths arising from AIDS related causes will not peak until 2020-2025 (United Nations, 2004).

In practice it can be very difficult for researchers to assess the true scale of the HIV epidemic upon individuals, their household and the community more widely due to a lack of data. In addition many of those living with HIV have negative experiences due to the social stigma associated with their condition. UNAIDS (2012) report that data from the People Living with HIV Stigma Index indicates over half are victims of verbal abuse due to their HIV status.[3] Educating people about how HIV can be spread is one of the first steps towards changing attitudes.

The number of people dying from AIDS related causes in the sub-Saharan Africa region is falling: down from 1.8 million in 2005 to 1.2 million in 2011 (UNAIDS, 2012). The number of new infections in the region is also declining, falling from 2.4 million new infections in 2001 to 1.8 million in 2011 which is a 25 percent decline in new infections over the period (UNAIDS, 2012).

[2] http://www.un.org/millenniumgoals/
[3] In Zambia, Rwanda and Kenya.

There has also been substantial progress in rolling out HIV testing within the region. Testing enables the medical profession to identify those who need antiretroviral therapy and to focus upon providing care to these people.

Antiretroviral therapy can lengthen the period of incubation between contracting HIV and developing AIDS (Kremer and Glennerster, 2004). According to recent estimates by UNAIDS in 2011, just over half of those living with HIV in sub-Saharan Africa were receiving antiretroviral therapy. South Africa and Malawi in particular have a relatively high uptake of treatment: both countries achieve a 60 percent coverage rate (UNAIDS, 2012). Moreover, in both Malawi and South Africa, steps have been taken to provide more resources for dealing with the HIV/AIDS epidemic. As HIV will continue to represent a challenge to health in the sub-Saharan Africa region, any discussion of health cannot ignore its presence.

There are several problems which are encountered within empirical research in the field of health and economic outcomes. One common problem is how we define health. I will discuss in the course of this thesis how health is not a simple one-dimensional concept. Some measures of health compare individuals against certain norms such as height and weight. Falling below critical cut-off points is then indicative of morbidity. Self-reported measures seek to elicit responses from individuals regarding their own perception of health. Ideally, if a range of indicators for health is available, then one should consider as many of these as possible within any empirical work.

Many of the factors that influence health outcomes such as education and where people live also influence other outcomes such as income. Therefore, a dual causality exists. To disentangle these dual pathways, research is required. However, even in the absence of conclusive agreement amongst researchers about how we examine these relationships, there does appear to be some common ground: recognition that some groups are more vulnerable than others and targeting these individuals should be a priority.

In order to be able to more readily identify those who are most vulnerable globally we need to collect better quality data. This includes collecting longitudinal data. Studying people over time enables us to better gauge the effects of morbidity and policy interventions upon their livelihoods. The better the quality of data that we have, the more informed we can be about the factors that make people vulnerable, and the more effective the interventions that can be put in place.

The ability of those who are most vulnerable to make informed decisions which are conducive to good health, is often constrained by a lack of knowledge and a lack of access. Within the child health literature there is a great deal of emphasis upon the education of the mother in particular. Education is generally found to improve childhood survival and other economic outcomes. What is less clear is whether other factors which lead to improvements in health outcomes also operate through education, and if so what these mechanisms are.

For adults, experiencing morbidity may commonly lead to a reduction in their working hours or withdrawal from the labour market altogether. If the health condition they experience is severe, they may never return to the labour market. Others within a household may also be forced to take time away from productive activities in order to provide care. Therefore, one consideration is to examine whether or not households experience large declines in their income, which could potentially impact upon their overall well-being, during periods of morbidity.

If morbidity leads to large reductions in labour supply and income, then one must also consider the ability of a household to adjust during times of hardship. If a household is able to make adjustments for a fall in labour supply and/or income relatively easily then the cost of adjustment is very low. The problem that a condition such as HIV/AIDS creates in sub-Saharan Africa is that many households may experience lengthy periods of illness by one or more household members who are often of working age. Following a loss of labour income, a loss of labour supply and/or an increase in healthcare expenditure, the ability of a household to maintain consumption during such a time may be considerably challenged, if the losses and/or expenditure incurred are very high. Severe morbidity can lead to households undertaking activities to minimise the adverse effect that morbidity has on them. Actions taken can include the use of savings, borrowing, selling assets or in some cases the use of insurance, to protect households against the fall in living standards they may experience. The problem arises when households are not able to adjust or there are large costs to adjustment, which is particularly challenging for poor households and may have implications for current and future welfare.

1.1 Study objectives

This thesis has three aims. The first part of the thesis uses household data from the Second Integrated Household Survey (IHS2) in Malawi, undertaken in 2004, to examine the determinants of child mortality in the country. The IHS2 has not previously been used to investigate child mortality within Malawi. The use of the IHS2 to examine this particular issue is therefore a novel contribution to the existing body of literature. The second part of the thesis will examine the effect that morbidity has upon the daily wage rate in Malawi. Previous research has considered mainly urban workers, whereas in Malawi the majority of individuals are located in rural areas. In addition, the examination of the effect of morbidity on casual (ganyu) labourers wages is a further contribution to the existing literature. The final aim of this thesis is to consider a range of different indicators of health in the form of "shocks" and the effect such shocks have upon household income in the KwaZulu-Natal Province in South Africa, using the KwaZulu-Natal Income Dynamics Study (KIDS) The KIDS dataset has not previously been used to investigate this particular question and therefore is a novel contribution to the literature. The specific objectives of thesis are:

i. To establish the factors which contribute to a reduction in child and infant mortality in Malawi. The loss of a child may be avoidable in many cases. To reduce the number of childhood deaths, we need to understand the causes of child death and the interventions which can have the greatest impact for improving child health. The literature identifies a range of factors which positively improve survival rates but relatively little work has considered Malawi. Given that we do not fully understand the role of social and cultural settings it is important to examine different countries as results from other countries may not necessarily be generalisable. In particular, the cause of death will be considered within the thesis whereas current literature does not make this distinction.

ii. To examine the extent to which morbidity reduces the daily wage that Malawian workers receive. There are a number of measures of health status considered within the literature. I will be using a measure of morbidity known as "disabled days" which represents the number of days someone was unable to work in a two week period due to poor health. I consider disabled days as an indicator of general morbidity, as this measure captures the incidence of illness in a specified time frame, rather than that directly associated with the wage received. If even a small amount of morbidity affects individual wages then one must consider the impact that this could have upon individuals and their households. In addition to examining wages, an examination of the effect of morbidity on ganyu rates which is a form of casual labour supply within Malawi is also undertaken. If reductions in wages are large and/or widespread then the ability of poorer households to cope could be quite problematic.

iii. To examine the impact that different indicators of "health shocks" have upon household income in the KwaZulu-Natal province. Households also directly report income losses associated with the shock experienced. This feature is most unusual and therefore allows me to make a comparison between directly reported income losses and estimated losses.

Examining different aspects of health, will aid our understanding of some of the key areas which policy makers should consider. The aim is to achieve sustainable improvements in the well-being of individuals, their household and the community more widely. Moreover, this thesis will seek to improve our knowledge about the importance of different factors which influence economic and health outcomes in Malawi and South Africa.

1.2 Outline of the thesis

The remainder of the thesis is as follows: In *Chapter 2* I consider the influential Grossman (1972) model of health. After outlining the model, a consideration of how health may be measured and the challenges associated with its measurement are explored.

The remainder of *Chapter 2* is split into three sections outlining specific literature associated with child mortality, wages and health shocks. The child mortality literature builds upon a framework which identifies the proximate determinants of child mortality. These are classified into five categories: maternal factors, environmental contaminations, nutrient deficiency, injury and personal illness control. Whilst others have added to this framework, these initial determinants remain prominent in all of the studies reviewed.

The second part of *Chapter 2* considers the human capital literature. Research has predominantly considered the effect of human capital returns such as education upon wages. This has been attributed to a lack of data on wages and productivity, particularly within developing countries. A number of studies which have looked at the impact of morbidity upon wages and productivity measures are reviewed, and there is a discussion of the problems encountered when conducting such analysis.

The final part of *Chapter 2* is an examination of a set of literature which considers the phenomenon of "health shocks". Within this field of literature, health shocks are defined in several different ways, including large healthcare expenditures, hospitalisation, ranking scales and large declines in weight. The discussion in this section also considers the different responses that households make in the presence of morbidity. The overwhelming evidence is that the cost of health shocks to household welfare can in some cases be mitigated. However, there is also evidence to suggest that the poor are particularly vulnerable and less able to cope with such adversity.

In *Chapter 3*, the survey data used in the subsequent analysis is presented. For clarity I have separated this chapter into two parts: the first part of the discussion is related to Malawi, and the second follows the same layout but with reference to South Africa. I begin the discussion for each country by presenting some background information about the country, before moving on to a more detailed discussion of the data sets used. In addition to describing the most important aspects of the dataset, I have also included for a discussion of how the data was collected. Certain sections of the datasets in both cases have either not been used at all or only a small number of indicators have been used.

Therefore, the depth of the discussion for some sections or indicators within the two surveys is very brief, as they are not particularly relevant for the purposes of this thesis

In *Chapter 4,* the patterns of childhood mortality within sub-Saharan Africa, and Malawi more specifically, are considered. The number and causes of infant and child deaths occurring in Malawi are examined using the IHS2. The majority of deaths involving children and infants can be associated to illness-related conditions; malaria is the primary cause of illness-related death in the region. I present a discussion of each of the factors identified within the literature as affecting child mortality and subsequently used within the analysis which follows. A probit and a Poisson model are used to estimate the effect that the commonly identified proximate determinants have upon child mortality rates in Malawi.

Chapter 5 explores the impact that morbidity has on wages in Malawi using Ordinary Least Squares (OLS) and Instrumental Variable (IV) estimation. This chapter uses a measure of morbidity known as "disabled days" to estimate the reduction in wages experienced by individuals due to morbidity, who have had at least one day off work in a two-week reference period,. Waged workers are selected as the wage for self-employed individuals can be difficult to clearly identify. Ganyu workers are also considered as an extension to the existing literature. The incidence and patterns of illness within the sample are presented to provide an overview of health conditions within Malawi.

Chapter 6 examines the income changes that households experience following a series of different "health shocks" which may have occurred between general elections. A general overview of the incidence of two health shocks: serious illness and injury or death, is presented. OLS and PSM estimation are utilised in order to estimate the potential effect of "health shocks" upon household income. A final step is an examination of the directly reported estimations of income losses associated with “health shocks”.

Chapter 7 concludes the thesis. It provides a summary of the main findings from each of the empirical chapters. Opportunities for further research and the policy implications of the findings are also considered.

Chapter 2: Adverse health in sub-Saharan Africa: a literature survey

There are two very broad components of health which researchers can consider; mortality and morbidity (Ghatak, 2010). Mosley and Chen (1984) suggest that whilst the majority of research focuses on mortality which can be readily observed and is a well-defined event; this can limit our scope for research. They indicate that this is particularly true when the sample size is small, as death can only happen once. Therefore, in order to examine mortality, you need either a large sample or to look at a sample over a lengthier time period. A particular concern which they raise is that in the social science field, we tend to concentrate exclusively on the end result, and thus we ignore the outcome of those who survive. The medical field in contrast, explores the other side, which is the outcome of those who have experienced morbidity.

McDowell (2006) indicates that the focus of measuring health in the last 150 years has moved away from mortality and instead now considers a broader focus such as that stated by the World Health Organisation (WHO) *"health, which is a state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity, is a fundamental human right…"*[4]

This chapter will proceed to consider why we demand health in *Section 2.1*. This is followed by *Section 2.2* which examines how we can measure health status. *Section 2.3* examines the field of literature specifically related to child morbidity, which is the focus of *Chapter 4*. *Section 2.4* examines the human capital theory literature which is the focus of *Chapter 5*. *Section 2.5* presents a discussion related to health shocks and *Section 2.6* concludes the chapter.

2.1 The demand for good health

Within the human capital literature (Becker, 1962; Becker 1975; Shultz, 1961; Ben-Porath, 1967 and Mincer, 1974), there are positive returns to having a greater stock of human capital, where human capital refers to the accumulation of skills and knowledge of an individual (Shultz, 1961). Investments in human capital are not free; they are

[4] p.ix.

costly, both in terms of time and possible direct costs such as the purchase of materials, fees and the opportunity costs associated with the investment such as forgone earnings. Therefore, if people are to undertake such investments for themselves, using their own resources, there needs to be a positive return to undertaking such investments.

Grossman (1999) notes that it is well established within the literature that people invest in themselves and in so doing, they increase their stock of human capital. Grossman (1972), however, suggests that there are important differences between health and other forms of human capital which are highlighted by Mushkin (1962). Mushkin (1962) suggests that education primarily affects the quality of the labour force, whereas health has both a quantitative and qualitative effect. Investment in health provides additional units of labour as well as a better quality of labour. Another important difference noted by Mushkin (1962) is that people invest in education on a pathway; one increment leads on to another and is a gradual process. In comparison, Mushkin (1962) suggests that investments in health are undertaken, effectively, to counteract what she refers to as “natural forces of biological selection”;[5] reducing the number of sickness days.

The Grossman (1972) model uses the framework of household production developed by Becker (1965) to consider the production of health capital. Within the Becker (1965) model households combine their time with market goods to produce commodities. These commodities could be things such as meals or leisure activities and they generate utility for the household members which they seek to maximise subject to a resource constraint. This constraint consists of the number of goods purchased within the market, the prices of these goods and sources of income. The Grossman (1972) model demonstrates that health will be demanded and produced as it has consumption and investment properties. As a consumption good, being healthy generates utility for people whilst sickness days represent disutility. As an investment good, a reduction in the number of sickness days enables people to have more healthy days. These healthy days can be dedicated to the production of both market and non-market goods which produce income and commodities which generate utility for the individual.

[5] p.135.

The intertemporal utility function of an individual is given in equation (1). The initial inherited stock of health is represented by (H_0), the current stock of health by (Hi), ($ø_i$) represents the service flow per unit stock of health, ($h_i = ø_i H_i$) is the total consumption of "health services" and (Z_i) is the consumption of all other goods and services (Grossman, 1972). The lifespan that someone will have is determined within the model and is therefore endogenous as death will occur when the stock of health falls below a critical level. The stock of health depreciates over time as we age and can be increased by undertaking investment in one's own health.

$$U = U(\emptyset_0 H_0 \dots, \emptyset_n H_{n,} Z_{0\,,} \dots Z_n) \qquad (1)$$

The individual undertakes investment in health by combining time (TH_i) with inputs such as medical care, appropriate nutrition and exercise/leisure (M_i) with human capital (E_i) which Grossman (1972) describes as one of the "environmental variables". Specifically the model highlights that the production of health is not exogenous; it is determined by the resources allocated to it and is influenced by other forms of human capital, such as education, which can both raise and lower the efficiency with which health is produced. The gross investment undertaken is determined by the production functions outlined in equation (2) for health and equation (3) for other commodities respectively. Where (X_i) refers to the goods input for the production of other commodities (Z_i) and (T_i) is time allocated to the production of the other commodities.

$$I_i = I_i(M_i, TH_i; E_i) \qquad (2)$$

$$Z_i = Z_i(X_i, T_i; E_i) \qquad (3)$$

In maximising one's utility, individuals face a budget constraint where both market goods and time are assumed to be scarce resources. Time can be used for working (TW_i), can be lost due to illness or injury (TL_i) which is a source of disutility or used for the production of health (TH_i) and other commodities (T_i) in any one period.

$$TW_i + TL_i + TH_i + T_i = \Omega \qquad (4)$$

The inclusion of lost time is important within the model to ensure that all possible uses of time (Ω) which may be thought of as the maximum number of days in a year are included. Moreover, lost time is inversely related to the stock of health; the greater the stock of health, the less time that will be lost through illness or injury. The distinction between lost time and health investment time in equation (4) is emphasised by Grossman (1972), as an increase in the amount of time invested in health increases the stock of health and also reduces the number of days lost due to sickness. By combining the time and goods budget constraints, the model shows that the optimal amount of investment and thus the optimal quantity of health capital can be determined.

In equilibrium the model shows that the present value of the marginal cost of investment must be equal to the present value of the marginal benefit of the investment and the discounted marginal benefit at age (*i)* will be equal to equation (5), where (G_i) represents the increase in the number of healthy days associated with a one unit increase in the stock of health. In order to convert this into a monetary value, Grossman (1972) uses the discounted wage rate which converts the increase in the time available for market and non-market activities into a monetary value and adds the discounted monetary value of the increase in utility associated with the increase in healthy time.

$$G_i \left[\frac{W_i}{(1+r)^i} + \frac{Uh_i}{\lambda} \right] \qquad (5)$$

Grossman (1972) points out that health capital is different to other forms of capital such as stocks or shares as such capital cannot be sold in a market and therefore implies that the gross investment cannot be negative. To increase the stock of health capital in the current period, one must increase the investment in health in the previous period. In so doing, investment in other assets will be lower and can be measured by the rate of interest forgone.

Using a pure investment model,[6] Grossman (1972) demonstrates that the optimal stock of health at any age will be determined by the intersection of supply and demand. Demand is given by the relationship between the rate of return on an investment

[6] Where healthy time does not enter the utility function.

(marginal efficiency of health capital, MEC) and the stock of health. The supply curve is perfectly inelastic as the cost of capital is independent of the stock of health. The intersection of supply and demand assumes that the marginal product of health is diminishing; additional healthy days produce diminishing marginal productivity.

In reality, the demand for health and thus investments in health will change over time due to a number of factors such as age, wage rates and other “environmental variables” which will now be considered. As people age, their stock of health will decline, we become physically and mentally less able compared to our younger selves. The Grossman (1972) model incorporates this natural aging process by increasing the rate of depreciation; in so doing, the cost of capital is increased which reduces the optimal stock of health. He concludes that if the demand for health is relatively inelastic (less than one), then as people age there will be an increase in gross investment to partially offset the fall in health stock as a result of the increase in the rate of depreciation. Gross investment and the rate of depreciation are positively correlated over an individual’s lifetime, whereas gross investment and the stock of health are negatively correlated.

A couple of additional issues are explored by Grossman (1972). The first considers the effect of rising wages. As wages rise, being healthy generates increasing returns to the individual as there are fewer days lost and more healthy days available. This causes the demand for health capital to also rise as it becomes increasingly more valuable. The second is related to the role of human capital. In particular, Grossman (1972) notes that more educated individuals are more efficient producers of earnings; they work faster and make fewer mistakes and are hence more productive ceteris paribus. If people are more efficient, then they effectively need to use fewer inputs to generate an output; more educated individuals need to undertake less gross investment. The result of this is that more educated individuals demand a greater stock of health capital.

The Grossman (1972) model concludes that health is a commodity; one which people desire as it generates utility both in terms of less sick time which in itself causes disutility but also due to the generation of healthy days which enable people to generate an income. At any given time, people determine the optimal stock of health capital that they need which will decline as people age given an increasing rate of depreciation. The demand we have for health is positively related to the wage we are paid; sick days are more costly to well-paid individuals and the educated will demand a greater stock of health.

2.2 Measuring health

If we are to measure health, then we need to be clear that the selected measure encapsulates what we mean by health according to McDowell (2006). Moreover, he states that when considering potential measurements for health, we need to consider what the specific indicator is measuring and what that measure tells us about population health status. In reality, the measurement of health may often be undertaken by using some form of self-rated health assessment, which is less costly than other methods (McDowell, 2006; Idler and Benyamini, 1997; van Doorslaer and Jones, 2003; Blane et al., 1996; Lee, 2015).

In the context of this thesis, the discussion that follows is particularly related to the measurement of physical health. The data utilised in *Chapters 4-6* has been collected during the completion of large household surveys which, as van Doorslaer and Jones, (2002) point out, very often have limited space for detailed questioning on health since health is one area of exploration amongst many. As a result, Rockwood (2015) suggests that there can be some disconnect between what the questions meant to respondents and what use this data then subsequently has. Despite this, the line of questioning in such surveys has become more standardised according to McDowell (2006). The advantage of this is the improved potential for comparisons across countries and over time which may provide useful insights into public health policy design.

Self-reported health is associated with an individual’s perception of their health status (Lee, 2015) rather than a clinical diagnosis. Economists have a different perspective to clinicians according to McDowell (2006); we wish to look at general aspects of health rather than specifics, and this is reflected within the line of questioning in household surveys. In addition, the popularity of self-reported and general measures of health has been linked to the findings that they can be reliable predictors of mortality (Franks et al., 1996; Idler and Benyamini, 1997).

Despite the aforementioned advantages and popularity of self-reported health measures, there is one important factor that needs to be highlighted; such measures require recall which may have important implications for the validity and reliability of the reported information. Recall relies on the individual's memory, when the event occurred, what the effect of that was, and so on. Loftus and Marburger (1983) suggest that recalled data may suffer from two problems; the event could be completely forgotten or people bring forward an event; remembering it sooner than it really occurred which they call "forward telescoping". They investigate this problem in a series of five experiments and conclude that using a landmark event[7] reduces the degree of reporting error that may arise when using recalled data.

Rockwood (2015) notes that time bounding where respondents are asked to indicate information only related to a specific period of time is a useful tool as it tends to enable respondents to provide a more specific response such as the number of times an event occurred rather than a vague estimation. However, the less common the event is, the longer the time period that one needs to consider, which then has to be balanced against the memory of the individual. An additional difficulty associated with using self-reported measures is the potential bias in reporting; reported health in such cases is then not reflective of their true health status (McDowell, 1996; Blane et al., 1996; Blaxter, 1989; Butler et al., 1989).

In terms of the actual questions asked within self-rated health surveys, respondents are often asked questions that can be quantified numerically or categorised with some form of index. Ilder and Benyamini (1997) review twenty-seven studies which considered self-rated health and demonstrate that the measures used include general health rating using a scale, the number of hospitalisations, the incidence of functional disability and chronic conditions, utilisation of health facilities and current weight and height. A comparison dimension may also be added so that respondents are asked to consider their health relative to a period in the past or to other people (Lee, 2015).

One of the difficulties associated with reporting health on a scale (which may also be associated with functional disability) is that in practice people may indicate that their

[7] Specifically, the eruption of Mount St Helens and New Year's Day were used.

health is better than what it may really be as they have “normalised” certain aspects of their health, which Blane et al. (1996) refers to as the *“iceberg”* phenomenon. Blaxter (1989) finds evidence that there are individuals whom the medical profession would consider were not in "good health" but who indicate that they are in good health. Health in this sense is more reflective of one’s perception of health as opposed to actual health according to Butler et al. (1987) Therefore, when an individual reports their health status it *may* be sensitive to where they place their baseline or "normal" state of health. For example Blane et al. (1996) indicate that working class people view the absence of disease as an indication of good health, whereas middle class people have higher expectations, and wish to not only be free of disease and impairment but also have a feeling of well-being.

The degree with which functional disability may be reported may also be influenced by an individual’s circumstances. Blane et al. (1996) suggest that someone engaged in a more physically challenging job may find it more difficult to cope with morbidity than someone in an office based job where they are sitting for prolonged periods of time, for example. Equally, someone with a long term employment contract may find that they can take time off work more readily than someone who is on a short term contract. Someone who has less job security may be more concerned about the possibility of losing their job by taking time off, for example. The potential risk of losing your job, may be weighed up against the cost of working whilst ill and therefore affect how people report and/or display morbidity (Blane et al., 1996).

McDowell (2006) indicates that measures of health which capture functional disability are numerous; over a hundred potential measures exist within the literature. Such measures can be narrow in terms of examining specific health conditions or broader to consider a more general measure of health. McDowell (2006) indicates that initial measures in this area examined impairment, then disability and finally the degree of handicap once it was realised that what really mattered was assessing the need individuals have for care.

In his review of physical health, McDowell (2006) considers seventeen different studies which use a range of different indices for physical disability; the inability or difficulty that one may experience when trying to carry out particular activities. A number of the scales are broadly similar, using a score to assess functional ability which could be recorded from observation or by asking the individual to indicate their ability to perform the activity in question. One of the problems highlighted by McDowell (2006) with a number of the scales reviewed is that relatively minor ailments are not picked up by the ADL (Assessment of Daily Living) type approaches; they are much more applicable to chronic health conditions and the elderly. Notably, many of the studies they reviewed were undertaken with elderly patients or those in rehabilitation of some kind.

One of the problems with many of these studies noted by McDowell (2006) is that despite attempts to standardise the questions and answers to develop comparability, very often there was some variation. Slight variations can at times be sufficient to impede comparability, which leads to a discussion of two important measures that seek to address this issue; QALY (Quality Adjusted Life Years) and DALY (Disability Adjusted Life Years).

As Grossman (1972) has established, health is demanded by people as it provides utility. QALYs and DALYs, which come under what Gold et al. (2002) refer to as HALYS (Health-adjusted life years), take this factor into account; it is not just about the quantity but also the quality of health that one has, which Torrance and Feeny (1989) suggest is advantageous.

The Quality-Adjusted Life Years measure was first developed in 1968 in a study of renal failure according to Drummond et al. (2005), although at that point the phrase QALY was not coined; by the mid-1970s, however, Zechauser and Shepard (1976) used the phrase in their exploration of how to value life. In the 1968 study by Herbert Klarman it was found that patients who underwent a transplant had a better quality of life than those who remained on dialysis (Drummond et al. 2005). The Klarman research showed that it wasn’t just about giving people greater longevity; it was also about how pleasant that additional time was. Having a long life expectancy may not be a positive experience for someone if much of that time was spent in pain or discomfort due to disability and disease. Following on from the research, the team calculated the cost of each year of life gained as a result of the different treatment options (Drummond et al. 2005), which is now a common application of the technique within the public health policy field (Gold et al., 2002).

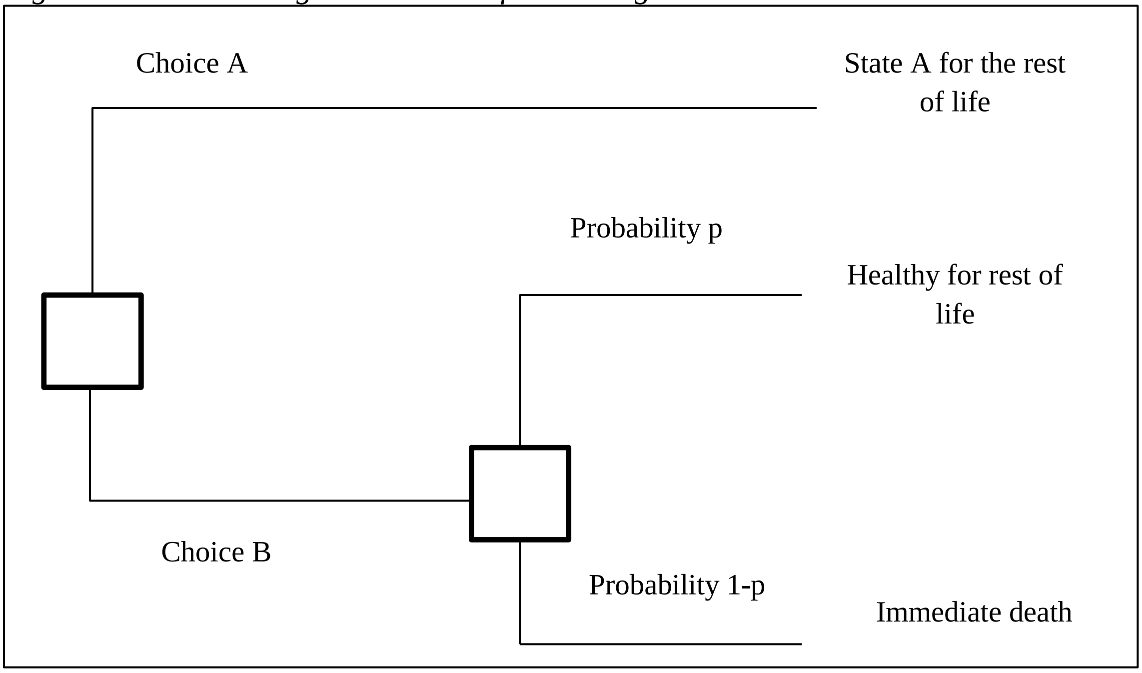
To calculate a QALY, weights which represent the quality of life experienced (a measure of utility or preference) are assigned. Perfect health takes on a value of one whilst death takes on a value of zero; the closer the value is to one, the more desirable is the state of health experienced (Gold et al., 2002) Within the development of QALYs, some states of health are assumed to be worse than death; an individual is in so much pain and discomfort that death is viewed as a better situation; such cases are assigned negative values to take account of this. In terms of valuing the utility derived from different health states there is a range of valuation methods which can be considered; standard gamble, time trade-off, rating scale, multi-attribute utility and person trade off (Weinstein et al., 2009).

The standard gamble method in the context of health presents possible outcomes to people; a certain situation and a gamble (due to uncertainty) where the outcome could be “better” but could also be worse. Respondents then provide an indication of the level of probability that would be needed for them to remain indifferent between the “better” state and the status quo (Torrance and Feeny, 1989).

An alternative measure is the time trade-off method, which Torrance and Feeny (1989) indicate gives similar results to the standard gamble method. With this method people compare two states again; living the rest of one’s life in less than perfect health or a shorter length of life in which you are healthy. The advantage of this particular method is that it is easier for people to identify with, according to Torrance and Feeny (1989).

The rating scale method uses visual interpretation; on a line based on the best and worst states of health, an individual places states of health along the continuum. Gold et al. (2002) describe this as being like a “feeling thermometer”. Weinstein et al. (2009), however, suggest that the rating scale method is not considered to be as robust as the standard gamble or time trade-off method but Gold et al. (2009) indicate that actually this approach is easier for people to understand.

Figure 2.2.1 Standard gamble method for eliciting utilities



Source: Reproduced from Torrance and Feeny (1989)

This therefore leads to an extension of the approaches in the form of what are referred to as multi-attribute measures which consider the range of health states across different attributes[8] such as; sensation, mobility, cognition, and so forth (Feeny et al., 1995). A final approach that can be considered to elicit utility values for health is the person trade off. Within this approach one might be asked to make decisions about how they might help groups within a population (Weinstein et al., 2009). Under this approach the respondent is making decisions for others rather than their own personal preference. As with all variations of any method; different methods of eliciting utilities lead to differences in the reported utility values (Gold et al., 2002).

Thus far, this part of the discussion has focussed on QALYs. I will now turn to another popular measure within the health economics literature which is DALYs. DALYs are an indicator for the burden of disease, measuring the loss of function people experience as a result of morbidity (Sassi, 2006). They were developed as a result of the 1993 partnership of the World Bank and the WHO for The Global Burden of Disease study, which gave rise to a series of very comprehensive reports.

[8] There are different measures which have a number of categories such as the Euro QOL 5 item scale; EQ-5D or the 6 item scale SF-6D, as two such examples.

This measure was designed to quantify the burden that disease and the subsequent debilitation such disease has on individuals (Gold et al., 2002). DALYs differ from QALYs as they also take into consideration the age of the individual and attach different weightings to the lifespan one has left at different ages. Young people and the elderly are given smaller weights as people of middle age may have caring responsibilities for both younger people and the elderly (Sassi, 2006; Gold et al., 2002; Drummond et al., 2005). A DALY can be used to measure the difference between what health is and an ideal (Gold et al., 2002) In contrast to QALYs where a value of one denotes the best state of health; zero is used to indicate the absence of disability in the case of a DALY and so lower rather than higher scores are preferable (Sassi, 2006; Gold et al., 2002). One of the difficulties that DALYs have associated with them is that in reality people may very often have a number of health conditions, particularly as they get older, but DALYs cannot separate out different conditions from one another (Gold et al., 2002). Moreover, one can also not separate out treatments from one another for different ailments which may have side effects in themselves according to Gold et al. (2002). Therefore, QALYs may be preferred to DALYs as the former considers health in a general sense rather than as the consequences of specific conditions (Gold et al., 2002).

As with QALYs, there are different ways in which information can be collated. Gold et al. (2002) suggest that health professionals are used to develop DALYs, particularly as they will have an informed knowledge of differing conditions and in the case of the World Bank and WHO study, nine expert groups were used to calculate the weights. The approach used is similar to the person trade-off previously discussed to consider differing degrees of disability and may consider physical impairment, cognitive ability, pain and also sexual/reproductive function (Gold et al., 2002). As in the case of QALYs people are then asked to weigh up different options for different groups of people based on different health interventions and ailments. One of the criticisms of this approach is that the weightings given to differing conditions are based on expert opinion rather than the actual experiences of real individuals who experience the condition. Not only that, but the impact on any one individual can be rather variable depending on their circumstances (Grosse et al. 2009).

QALYs and DALYs represent measures of health that are standardised[9] and are therefore an important part of the health policy landscape when considering the cost-effectiveness of differing interventions within populations. The next part of the discussion considers some alternative measures that can be used to infer the health status of a population, before moving on to a discussion of the specific literature associated with the empirical work undertaken in *Chapters 4-6.*

One set of alternative measures put forward as being less subjective than self-reported measures of health and not requiring detailed information on different health states and calculations of associated utility weightings are anthropometric indicators; commonly height and weight or body mass index (BMI). These measures are argued to be less subjective as they can capture both past and current changes in health (Thomas and Strauss, 1997). These measures are more commonly observed for children to assess the presence and depth of malnutrition in developing countries. Such measures are also not viewed as being particularly invasive and can be used for large groups of people (Sicotte et al., 2010). One of the major drawbacks, however, is that the personnel carrying out the measurements need to be properly trained and so it is important that measurement error does not arise (Sicotte et al., 2010). There are also some additional anthropometric measures that researchers can use such as arm circumference, waist circumference and tricep skinfold (Sicotte et al., 2010).

Height in general reflects prior investments in health, particularly in early childhood and may also be associated with non-health human capital investments according to Thomas and Strauss (1997). Weight is also commonly used as a measure of health but one must remember that someone who is short and weighs very little may not necessarily be unhealthy. Equally someone who is heavy but very tall may also not be automatically considered as unhealthy (Thomas and Strauss, 1997).

Therefore, weight or height alone may not be particularly good indicators of health as there are other things to consider such as general build. This particular issue has led to nutritionists to use another measure which is body mass index (BMI). BMI is calculated

[9] That being said, there is a range of different weightings that can be applied which will lead to differences in results.

by taking someone's weight (normally in kilograms) and dividing this by the square of their height in metres. This calculation then gives a numerical value which is then compared to a chart which splits the scores up into categories; underweight, normal, overweight and obese. Critics of this measure suggest that it does not account for normal healthy variations between individuals, such as build and whether or not someone is particularly muscular, which would increase weight, and therefore potentially BMI.

However, in most household surveys, information is not collected on the BMI of adults. In fact, both of the household surveys I use for my empirical analysis do not include such indicators for adults. Therefore, it is not possible to use BMI as an indicator of health status in the subsequent empirical work that I undertake. However, there is recognition that such data could be helpful. *"Anthropometric data in adults are not currently available on a global or regional basis in order to compile meaningful and representative databases".* [10]

Another additional approach suggested by Thomas and Strauss (1997) is the use of nutrient intakes. Specifically, they consider the use of calorie and protein intakes. An examination of intakes can be used to assess whether people have sufficient intakes to ensure that they can carry out activities. Alternatively clinical measures can be examined to determine whether nutrient intakes were adequate (Webb, 2007). For example a lack of iron may present a clinical condition such as anaemia; a lack of vitamin A may be associated with eye conditions (McCollum, 1939).

Ferro-Luzzi (2003) reviews the methods used to evaluate individual food intake survey techniques but suggests that there are *"methodological issues: accuracy and representativeness [sic] of the results as well as logistics and cost of conducting the surveys"* Ideally, Thomas and Strauss (1997) suggest that each individual's intake is preferable, but this is quite intrusive and is therefore not very practical. Furthermore, Ferro-Luzzi (2003) highlights that of the 84 studies that were reviewed by Thompson et al (1997) over half were undertaken in North American and two-fifths were undertaken in Europe. Ferro-Luzzi (2003) alludes to the fact that the absence of using nutritional

[10] http://www.fao.org/docrep/005/Y4249E/y4249e0b.htm

intakes as an indicator of health in developing countries is perhaps indicative of its lack of suitability in such environments.

From a theoretical point of view an indicator of health that enables different studies to be compared has distinct advantages; differing experiences of interventions can be explored which can provide direction for other cases. The development of QALYs and DALYs has certainly been a step forward in this regard. However, both have substantial data requirements both in terms of reliability and scope. Detailed information on different diseases and conditions, by age and gender with specifically linked functional ability information, is required for the calculation of such measures.

Self-rated measures of general health bypass such significant data needs and there is a considerable amount of evidence to suggest that self-rated measures of health are a reasonable indicator of actual health (Butler et al., 1987). Moreover, a really interesting point raised by Butler et al. (1987) is whether alternative measures, which are perceived to be “better” measures of health as they are more objective, really are better. Assuming that objective measures are “better” is a strong assumption and one that may not necessarily be true according to Butler et al. (1987).

In the absence of the data requirements which are necessary to consider QALYs or DALYs, this thesis will use self-reported information; in the case of *Chapter 4* this information is mortality data. The concern with such data is reporting error in the absence of formal records. *Chapter 5* will utilise self-reported functional disability data with reference to the withdrawal from normal activities as opposed to a measure such as walking up a set of stairs, for example. I do not feel that the ability of an individual to carry out such specific tasks such as walking a particular distance relevant within the context of the chapter; the chapter explores the effect of morbidity on wages. I therefore consider a more general measure of morbidity to be appropriate in this case. *Chapter 6* explores the role of health shocks and the impact it may have on household income. Given that “shocks” may be highly varied, I again consider a broader measure of morbidity; those associated with illness or injury which cause someone to stop their normal activities.

I will now turn to some more specific literature, which relates to the three major empirical chapters contained within this thesis; determinants of child mortality and the effect of morbidity on individual wages in Malawi which is presented in *Chapters 4* and *5*. This discussion is followed by an examination of the impact that health shocks have on household income in the KwaZulu-Natal province in South Africa which is presented in *Chapter 6*.

2.3 A framework for examining population health

An important contribution to the literature examining child mortality is the Mosley and Chen (1984) framework. The purpose of their framework was to bring together some differing but key research ideas, which they believed needed to be examined collectively. Their belief was that the determinants of child mortality, whether they are economic or social, all worked through the same mechanisms, which they called proximate determinants. Using this method, they believed that researchers would be more able to identify the determinants affecting child mortality and thus policy makers could make more informed decisions, about the types of interventions that should be made, to reduce child mortality. As we have seen in the demand for health capital literature developed by Grossman (1972), health status is determined by some initial inherited stock of health. Therefore child health is important as it represents the initial endowment of health capital that will take people into adulthood, and this has potentially important consequences on their aggregate human capital stock in adulthood.

The Mosley and Chen (1984) framework is important within the child mortality literature as it sought to bring together the different fields of research; social science and medical research. Mosley and Chen (1984) note that previously social researchers focused on the relationship between social and economic factors and the influence that these have upon mortality. In contrast, the medical profession focused upon the process of disease, rather than mortality as a topic in itself. This approach, they argued, led to a gap in knowledge and understanding about child mortality. Hence, there was an opportunity to bring together these two fields of research.

Within the Mosley and Chen (1984) model there are five identified proximate determinants; these are factors which *"directly influence the risk of morbidity and mortality".*[11] The proximate determinants are grouped into five categories:

- Maternal factors
- Environmental contaminations
- Nutrient deficiency
- Injury
- Personal illness control

Mosley and Chen (1984) indicate that the first four factors influence the rate at which someone may experience morbidity in the first instance. Once sick, people can either recover or experience long-term consequences in the form of growth faltering and/or mortality. Socioeconomic determinants affect personal illness control and thus affect how we seek to prevent illness and/or treat it, which in turn, determines whether we remain healthy or sick. Mosley and Chen (1984) group the socioeconomic determinants into three broad categories; individual level variables, household level variables and community level variables.

The Mosley and Chen (1984) framework has been important, as Macassa et al. (2011) indicate that it is this framework, which has been fundamental to the development of the Demographic and Health Surveys, and which has enabled researchers to explore the determinants of child mortality across the globe. One of the key benefits of such surveys is that we can track the evolution of health over time and across countries, particularly with the development of more standardised measures such as QALYs and DALYs. This enables policy makers to be able to evaluate the effectiveness of different interventions in a range of different environments.

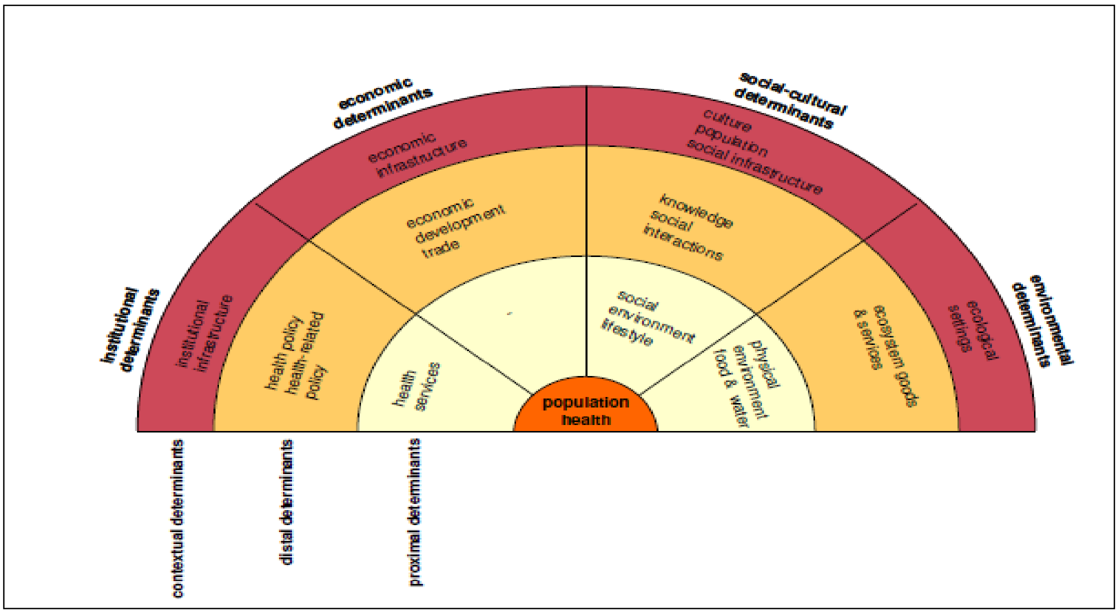
The benefit of using the proximate determinants framework is that by identifying the key risk factors for child survival, policy makers have been able to develop targeted programmes aimed at reducing these risk factors in differing contexts. However, Macassa et al. (2011) argue that a broader approach needs to be considered, particularly

[11] p.27.

in the sub-Saharan Africa context, with additional acknowledgement of the social factors that affect child mortality.

A framework that seeks to address these concerns is presented by Huynen et al. (2005) who build upon the original Mosley and Chen (1984) framework to take into account what are referred to as distal and contextual factors. By broadening the scope of the framework, one can take into account the relationships that interplay with one another and therefore build a more encompassing model of population health. This model is far more complex than the framework developed by Mosley and Chen (1984), as shown in Figure 2.3.1.

Figure 2.3.1 Multi-nature and multi-level framework for population health



Source: Huynen et al. (2005)

What this model highlights is that the proximate determinants are not the only areas of policy that we need to think about. It is not enough to know that the local environment can have positive or negative effects on health outcomes; we also need to have an understanding of how people live together in communities in order to understand how disease may spread; we also need to understand how knowledge is exchanged as Blane et al. (1996) clearly show that the relationships we have with one another can affect the way in which we respond to different health situations; which would serve to reinforce or undermine health care interventions.

By reflecting on health more broadly we can consider not only what exogenous variables influence health but also the mechanisms through which they might work; via the inclusion of mediating variables. For example, education and income are considered as independent variables in *Chapter 4* with child mortality as the dependent variable. Education and income are however closely linked as education has a causal effect on income (Deaton, 2003). Income can therefore be viewed as a mediating variable for the impact education has on child mortality.[12]

The evidence is far from conclusive in this area. Notably, Deaton (2003) highlights that there is a great deal of scepticism surrounding the link between income to health and that more recently, it is believed that, causality runs from *"good health to higher earnings, and from poor health to the inability to work and, in some countries, to running down assets to pay for care"*[13]. Therefore, Deaton (2003) suggests that there is "*some effect running from income to health, a position that has increasingly gained currency among economists in recent years"*.[14] The Grossman (1972) model shows us is that health is used to generate income via additional healthy days. In the next period this income can be used to purchase goods and services such as; food, housing, medical and public health services, education, leisure and health-related research (Preston, 2007) which positively contribute to the production of health capital and the avoidance of morbidity.

One of the main advantages to the Mosley and Chen (1984) framework is that in terms of data requirements it is relatively straightforward. Information on maternal characteristics such as age, education and birth intervals are captured within the DHS (Demographic Health Survey) and are often recorded within other types of household surveys. The IHS2 which is utilised for the purposes of *Chapter 4* includes similar data with the exception of birth intervals. The omission of this particular covariate is problematic and therefore an alternative, which is the number of births that have occurred in the last two years, is used as a proxy instead in an attempt to overcome this problem. The DHS and household surveys also include details on some of the

[12] In *Chapter 4* I find that it is the direct impact of education that matters rather than the mediator (income).
[13] p.118-119
[14] p.119.

environmental factors that may place individuals at an increased risk of morbidity; the IHS2 includes information on factors such as the water source, sanitation and rubbish disposal at the household and is also included within the KIDS data which is used in *Chapter 6.*

The IHS2 data unfortunately does not have information on the third area of proximate determinants; nutrient deficiency. However, there is some (albeit limited) information which can be used to assess potential longitudinal nutritional deficiency in children and this is subsequently explored within *Chapter 4.* The final area for inclusion with the proximate determinants framework is personal illness control; here data is required that will capture the prevention of disease. This can be simple activities such as hand washing to minimise the spread of disease generally or a more direct and targeted intervention such as the use of vaccinations and bed nets to prevent a condition such as malaria. The IHS2 again reports information on some areas of personal illness control and an important finding from *Chapter 4* is the evidence that bed nets have a very large and positive effect on reducing child mortality in Malawi. Another important set of indicators which appear within both the IHS2 and the KIDS data is information on the local healthcare provision in terms of the number of medical personnel, and the distance one may have to travel to reach such assistance. Data on personal illness control is an important factor within all three of the empirical chapters and is therefore utilised within the models that are subsequently developed.

The more advanced framework of Huynen et al. (2005) points to the interplay between differing groups' behaviours, the influence of policy and other cultural factors. The reality, though, is that capturing this is very difficult. In particular, one of the issues is that we don't know what would have happened in the absence of some kind of policy intervention; would the outcome be similar without the intervention or very different; we just don't know what the counterfactual would be, which is a very challenging issue within the social sciences. In practice, as is the case within this thesis and indeed in much of the empirical evidence I will go onto review, a limitation to our knowledge and therefore understanding lies within the data we can reasonably collect and use for modelling purposes.

2.3.1. Child health and child mortality patterns

The World Health Organisation (WHO) uses a range of measures to examine mortality and health. There are some key measures for examining child mortality and this thesis will specifically consider under-fives and infants. Child mortality is defined as the *"probability of a child dying before it reaches the age of 5 years".*[15] Infant mortality is *"the probability of dying by the age of one year".*[16] Reducing child mortality[17] is one of the Millennium Development Goals (MDG's). The goal is to *"reduce by two-thirds, between 1990 and 2015, the under-five mortality rate".*[18]

Children are more vulnerable to health risks than adults as their immune system has had less time to develop. Overall there is progress in reducing the rate of child mortality, with a recent estimate from CME Info which is a partnership between the WHO, UNICEF and the World Bank, indicating that the number of childhood deaths has declined from almost 12 million in 1990 to 6.9 million in 2011.[19] Progress is also becoming faster, with the reduction in childhood deaths rising from 1.8% per annum between 1990-2000 to 3.2% between 2000-2011.[20] However, despite these positive gains the majority of childhood deaths are still concentrated in the sub-Saharan Africa region, with one in nine children from the region dying before they reach five years of age.[21]

If we can make significant progress in other countries, then we need to understand what is happening in sub-Saharan Africa and how we can make gains in reducing child mortality here as well. As the Mosley and Chen (1984) and its extensions show us, children need good nutrition; disease prevention (in the form of vaccinations) and a clean environment to not only survive, but also to flourish. Children cannot take care of themselves and so it is important that we understand how child health and therefore

[15] http://www.who.int/healthinfo/statistics/mortality_child/en/index.html
[16] http://www.who.int/healthinfo/statistics/mortality_child/en/index.html
[17] This is MDG 4.
[18] http://www.who.int/topics/millennium_development_goals/child_mortality/en/
[19] http://www.childmortality.org/
[20] http://www.childmortality.org/
[21] http://www.childinfo.org/mortality.html

child survival can be affected. Policy actions are then undertaken to ensure that we put a safety net to catch these children, if such provisions are not readily available.

Progress is being made in terms of data collection with the introduction of the Inter-agency group for Child Mortality Estimation (IGME). This group was formed in 2004 and is a partnership between UNICEF, WHO and the World Bank as well as the United Nations Population Division of the Department of Economic and Social Affairs. This group is seeking to compile all the data that is currently available on child mortality, from different types of surveys within each of the countries, in order to gain a better view on what is happening in each case. As noted previously, it would be desirable to have a broader view of the interactions between different aspects of health which is what the Huynen et al. (2005) framework seeks to incorporate. However, in order to do this better data is required; both in terms of scope and across different environments.

A particular challenge within the sub-Saharan African contest is collecting reliable data on births, deaths, survival time as well as a wealth of information about the household characteristics which are used to develop a model so that researchers can identify the differences that exist between groups. Houweling and Kunst (2009) note that very often in the low and middle income countries such detailed data is very difficult to come by, as the registration system for all this information is underdeveloped. Added to this is the fact that the most vulnerable people; the poor and least educated, are even less likely to submit this information (Houweling and Kunst, 2009) or there may be reporting error due to issues associated with recall. The Demographic and Health surveys are one of the key datasets which are currently used for examining child mortality and have been in place since the mid-1980s, with a relatively large coverage of countries and a series of questions pertinent to identifying the influences upon child mortality (Houweling and Kunst, 2009).

However, datasets such as the IHS2 also have coverage on a number of the key areas that are pertinent to examining child mortality and also have additional data that is worthy of consideration. In particular, the IHS2 data which is reviewed in more detail in *Chapter 3* includes information on household income. Given the ongoing debate and lack of clear agreement on the role of income within the health literature, exploration of this topic should not be ignored. By utilising a different data set such as the IHS2 within this thesis I am able to explore whether income does matter in the case of Malawi, which is a factor not explored within the Manda (1999) study. A recent report by the IGME group notes that there has been substantial progress in reducing child mortality across many parts of the world. Since 1990 worldwide child mortality has fallen by 41 percent (87 deaths per 1,000 live births). This reduction is even larger in some regions; rates have fallen by over 50 percent in Eastern Asia, Northern Africa, Latin America and the Caribbean, South-East and Western Asia (UNICEF, 2012) as shown in Table 2.3.1.

However, whilst some regions have made significant progress, there are some regions such as Oceania, sub-Saharan Africa, Caucasus and Central and Southern Asia where progress has been much slower. Again this is where a broader and inclusive framework such as that suggested by Huynen et al. (2005) has its merits. With appropriate data it may be possible to identify the differing contextual and distal factors that are pertinent in such cases. The question is that if we can achieve considerable gains in some areas why can this not be replicated in others, and what might the missing link be in such cases. It is anticipated that these regions will not meet the target of the MDG 4, relating to the target reductions in child mortality (UNICEF, 2012). In fact, around half of all the childhood[22] deaths take place in only five countries; India, Nigeria, Democratic Republic of Congo, Pakistan and China. Nigeria and India account for over a third of all childhood deaths (UNICEF, 2012).

Worldwide, the leading causes of death in children under five are pneumonia (18% of all under-fives deaths); pre-term birth complications (14%); diarrhoea (11%); birth related complications (9%) and malaria (7%). Moreover, over a third of all deaths can be attributed to malnutrition in some way (UNICEF, 2012). What is most saddening is that these children are dying from *"preventable causes and treatable diseases even though the knowledge and technologies for life-saving interventions are available"*[23] (UNICEF, 2012).

[22] Under five years of age.
[23] p.2.

UNICEF (2012) indicates that reducing child mortality is possible with the acceleration of preventative and curative policy interventions to tackle the main causes of childhood death. Sub-Saharan Africa and Southern Asia are the key targets for this expansion and are a priority. Expanding the coverage of health interventions is particularly important for helping the poor to access basic services and ensuring that supplies get through to people. As the Grossman (1972) model shows us, investments in health provide us with additional healthy days and the efficiency with which we may utilise such investments is determined at least in part by another area of human capital; education. It is not enough to just provide inputs to people, we need to think about how such inputs are then used in order to generate the greatest gains in outcomes.

According to figures published by UNICEF (2012), Malawi's under-five mortality rate per 1,000 live births was 227 in 1990. This had fallen to 164 in 2000 and 83 in 2011. In addition to improvements in other areas of the proximate determinants which are key components of development policy such as access to clean water and sanitation, vaccination coverage and improvements in education, direct intervention also needs to be undertaken to provide a safe environment at birth. This includes the encouragement of exclusive breastfeeding, cord care and postnatal care for mothers according to UNICEF (2012). This very early intervention has significant effects as it *"brings a triple return on investment, saving mothers, newborns and unborn children".*[24] UNICEF (2012) suggest that some of these interventions are relatively low cost and it is therefore a scaling up exercise, to improve premature birth conditions, which could, reduce neonatal deaths by three-quarters (UNICEF, 2012).

2.3.2 Empirical evidence on the determinants of child mortality

In this section, I will review some of the literature that has examined the determinants of child mortality, across a range of countries. The aim of this is to observe the key factors that have been highlighted within the existing literature and the approaches that have been undertaken. This review is not intended to be exhaustive, but to provide an overview of the current literature. As noted within the introduction to this section, the proximate determinants are more easily identified within empirical work. Consequently,

[24] p.13.

the majority of the findings which I refer to are also associated with the effect that these proximate determinants have on child health outcomes rather than the preferred case which would capture the distal and contextual determinants. The latter would be a recommendation of Macassa et al. (2011) who imply that the Mosley and Chen (1984) framework is too simple a framework for capturing the complexity of child survival.

Houweling and Kunst (2009), in their global review of studies on child mortality, find evidence to indicate that the probability of childhood death is strongly correlated to the socio-economic status of parents and/or the status of the household where the child is born. The risk of a child dying is far greater in a poor household and one where the mother is relatively less educated (Houweling and Kunst, 2009). This is repeatedly found across the range of studies they review.

An important area of exploration which is followed within this thesis is the potential role of income upon child health outcomes. Manda (1999) specifically considers Malawi but crucially does not consider income as an additional explanatory variable as his study uses the DHS. Therefore *Chapter 4* seeks to build on this particular study by exploring income as an additional variable. Malawi is also an interesting case given the prevalence of HIV in the country. As the Huynen at al. (2005) framework clearly highlights, there may be important contextual issues that need to be considered and it would seem that countries such as Malawi with high rates of HIV may warrant additional examination. As Barnett and Whiteside (2006) highlight, *"HIV/AIDS is not the first global epidemic, and it certainly won't be the last: it is a disease that is changing human history".*[25] Whilst it is not possible due to a lack of available data within the IHS2 to specifically explore the potential role of HIV within Malawi, it is certainly a factor to be aware of.

This thesis seeks to explore whether there is a direct role of income given that the IHS2 has available data and the fact that the previous study for Malawi did not include it and indeed nor do a number of other studies due to a lack of available data. Specifically, *Chapter 4* seeks to explore whether income per se is a driving factor behind improvements in health and survival rates of children. Case et al. (2002) find evidence

[25] p.27

that household income does have a positive effect on children's health status in the United States; which they indicate is a result of such households being able to slow the rate of health depreciation. One must not forget, though, that the relative income position of US households is hugely different to that of a country within sub-Saharan Africa such as Malawi. Ssewanyana and Younger (2005), who do consider a low income country (Uganda), also find that increased household income reduces infant mortality. Notably, however, they use the DHS and so are unable to present a more direct indicator of income, whereas the IHS2 does enable me to consider income rather than a proxy.

There is some evidence presented by Houweling and Kunst (2009) to support the view that child mortality is higher amongst households where the household is of a lower socio-economic grouping. However, within such households, the proximate determinants that Mosley and Chen (1984) originally described are also less favourable; greater exposure to malnutrition, greater incidence of disease and poorer maternal and healthcare related characteristics. Wealthier households, in contrast, are able to invest in health and in so doing raise the stock of health capital (Case et al., 2002).

The question, therefore, explored with the IHS2 data for Malawi is, whether lower income households experience more mortality. Effectively, is it income or, as the framework developed by Huynen et al. (2005) suggests, is it factors that are associated with low incomes that are important. Moreover, if the evidence from the IHS2 data suggests improvements in income are relevant, what would be the magnitude of gains required to make realisable differences to the child mortality rate within the country. SSewanyana and Younger (2002) find evidence that income does make a difference but they also indicate that the effect is very small. Consequently they argue that even with continued income growth, such gains in terms of reducing the infant mortality rate would be small. Therefore, one must consider that even if it is found that income is important, the relative magnitude of such gains needs to be considered. Moreover, given the range of potential policy interventions that can be made, one must prioritise which interventions to make to improve livelihoods; some of which may present greater gains to households at quite low levels and be relatively simple to implement.

The role of education

The importance of the effect that education has upon child mortality has been developed on the foundations of a seminal paper by Caldwell (1979), who examined the effect of maternal education on child health outcomes in Nigeria. He also notes the positive benefits that education has upon child survival rates in the Philippines, Costa Rica, Mexico and Bangladesh. The importance of this paper is that it showed that even after controlling for other factors including control for the education and occupation of the father as well as socio-economic status, it was the mother's education which was found to have *"an independent, strong and positive impact on the survival of children"*[26] (Caldwell, 1994). Caldwell (1979) suggested that there were a number of ways in which maternal education could improve the probability of child survival. These were; more educated mothers break from traditional practices and are less accepting of illness. Therefore they try new things and are more proactive in their approach. More educated women are also argued to be more assertive with medical staff and command their attention. Caldwell (1979) indicates that this is something uneducated women may really struggle with as they are likely to be ignored. The last effect presented is suggested to be perhaps the most important; education alters the relative position of the mother within a household/family dynamics.

Education has continued to be found as an important factor in reducing the probability of child mortality in a range of studies. Bicego and Boerma (1993) examine child mortality rates in seventeen developing countries. They find that mortality rates in very young children (aged 0-23 months) are much higher amongst those with mothers who possess little education. Furthermore, even at very low levels of education, Basu and Stephenson (2005) find evidence that maternal education has a significant effect on child survival prospects. Basu and Stephenson (2005) suggest that even a small increase in education can be a *"magic bullet"*[27] for improving lives. An important issue to note at this juncture is that within the IHS2 a considerable proportion of women indicate that they have not completed formal education to primary school level. Therefore, as Basu and Stephenson (2005) suggest there may be considerable gains to improving education within Malawi which is also part of the longer-term development policy (MDG4 which

[26] p.224.
[27] p.2011.

is associated with improvements in education for young girls), given that there are inter-generational effects; well educated women make more informed choices which trickles down through each generation of children.

Despite the finding that education in general has a positive effect in reducing child mortality rates; there is a lack of clarity regarding the pathway along which education operates. Hobcraft (1993) concludes that studies which have attempted to look at this specific issue have been inconclusive and present contradictions. That being said, he proposes that the pathway between education and survival rates may differ across the reviewed studies as there are unaccounted for differences in the cultural settings which again takes us back to the argument put forward by Huynen et al. (2005) for using a multi-level framework for population health; modelling the cultural setting is beyond the scope of this thesis. Therefore, *Chapter 4* will only explore the potential gains in reducing child mortality within Malawi as a result of increasing education.

Whilst much of the research focuses on the mother's education, Caldwell (1994) and Hobcraft (1993) indicate that the education of a father should not be ignored. In order to explore whether the role of father’s education is relevant in the Malawian context, ideally an indicator for paternal education would be included within the estimations which follow in *Chapter 4.* However, the identification of the parents is at times very challenging with the IHS2 (a full discussion of this issue is presented in *Chapter 4).* Therefore this particular matter could not be explored. This is where the DHS data is advantageous as Manda (1999) is able to include measures of paternal education and finds evidence that the father’s education matters, as it is an additional indicator for the socioeconomic status of the household. The omission of paternal education therefore represents a potential omitted variable bias within the work that follows; the cost of using the IHS2 is the inability to include some well documented proximate determinants at the gain of exploring the potential role of income.

Putting aside some of the debate about how education operates, Hobcraft (1993) suggests that improved education leads to mothers seeking better health services for their children which I explore in the next section. Better educated mothers have an improved knowledge about disease and are therefore more likely to take sick children to seek medical advice. Blane et al. (1996) suggest, how we perceive our health affects when people may seek medical assistance. If people are more aware of their health status (illness), then the speed with which people seek assistance will be quicker and thus the likelihood of recovery may be somewhat greater. In addition, Hobcraft (1993) finds that children of more educated mothers are more likely to have received a full vaccination programme.

Mothers who are more educated may also make the decision to have fewer children and focus the resources they have on improving the quality of life for those children and thus reduce the probability of mortality. This may be in terms of nutrition and human capital investments (Hobcraft, 1993). Whilst Hobcraft (1993) notes that there is not a great deal of evidence to support this view, he indicates that the limited number of studies which do examine these issues, do tend to support the proposition that better educated mothers, provide additional support for their children. This translates into improved survival rates.

A final pathway identified by Hobcraft (1993) for the transmission of education to reductions in child mortality is via personal empowerment which is outlined by Cleland (1990). Hobcraft (1993) describes three components of empowerment identified by Cleland (1990) which is similar to some of the ideas proposed by Caldwell (1979). These are; instrumentality, social identification and confidence. The suggestion is that, if women are more educated, they are more able to make decisions regarding their own use of health services and therefore by default, to make these decisions for their children.

Maternal characteristics

In addition to maternal education there are other important maternal characteristics that need to be considered. Maternal age, birth spacing and breastfeeding are three factors which warrant consideration. Pebley and Stupp (1987) indicate that there is a J- or U-shaped relationship which exists between the age of the mother and the risk of child mortality. The risk of child mortality is greatest at the two extremes; older and very young mothers.

Table 2.3.1 Levels and trends in the under-five mortality rate, by Millennium Development Goal region, 1990–2011 (deaths per 1,000 live births)

Region	1990	1995	2000	2005	2010	2011	MDG target 2015	Decline (%) 1990-2011	Annual rate of reduction 1990-2011	1990-2000	2000-2011	Progress towards MDG target 2011*
Developed regions	**15**	11	10	8	7	**7**	**5**	55	3.8	4.2	3.5	On track
Developing regions	**97**	91	80	69	59	**57**	**32**	41	2.5	1.9	3.1	Insufficient progress
Northern Africa	**77**	59	45	34	26	**25**	**26**	68	5.5	5.4	5.5	On track
Sub-Saharan Africa	**178**	170	154	133	112	**109**	**59**	39	2.3	1.5	3.1	Insufficient progress
Latin America and the Caribbean	**53**	43	34	26	22	**19**	**18**	64	4.8	4.4	5.2	On track
Caucasus and Central Asia	**76**	70	61	52	44	**42**	**25**	44	2.8	2.2	3.3	Insufficient progress
Eastern Asia	**48**	45	35	24	16	**15**	**16**	70	5.7	3.3	7.8	On track
Excluding China	**28**	36	30	19	17	**17**	**9**	38	2.3	-0.7	5.0	On track
Southern Asia	**116**	102	88	74	63	**61**	**39**	47	3.1	2.8	3.3	Insufficient progress
Excluding India	**119**	103	87	72	62	**60**	**40**	50	3.3	3.2	3.4	Insufficient progress
South-eastern Asia	**69**	57	47	37	30	**29**	**23**	58	4.1	3.9	4.4	On track
Western Asia	**63**	52	42	37	21	**30**	**21**	52	3.5	4.1	3.0	On track
Oceania	**74**	67	61	56	51	**50**	**25**	33	1.9	1.8	1.9	Insufficient progress
World	**87**	82	73	63	53	**51**	**29**	41	2.5	1.8	3.2	Insufficient progress

Source: UNICEF (2012)

Note: All calculations are based on unrounded numbers.
a "On track" indicates that under-five mortality is less than 40 deaths per 1,000 live births in 2011 or that the annual rate of reduction is at least 4 percent over 1990–2011;"insufficient progress" indicates that under-five mortality is at least 40 deaths per 1,000 live births in 2011 and that the annual rate of reduction is at least 1 percent but less than 4 percent over 1990–2011. These standards may differ from those in other publications by Inter-agency Group for Child Mortality Estimation members.

The risk of child mortality born to older mothers is associated with the fact that an older mother's reproductive system is not as effective as that of a younger woman, as it is tired and worn out (Manda, 1999). At the other end of the spectrum, a very young woman may not have fully developed her reproductive system and may have received less care during their pregnancy (Manda, 1999).

A really important relationship that must be considered is that of fertility and mortality; Rosenzweig and Shultz (1983) indicate that there is a strong correlation between birth rates and the rate of infant mortality across a range of different settings. It is proposed that there are several possible explanations for this. One suggestion is that breastfeeding plays an important role in increasing birth spacing as it helps to reduce fertility (Rosenzweig and Shultz, 1983; Manda, 1999); fewer births within a given period of time reduce the strain put on the mother's reproductive system and health more generally, which can lead to a reduction in the likelihood of premature birth and or weaker children (Pebley and Strupp 1987; Rosenzweig and Shultz, 1983; Manda, 1999). Moreover, breast milk does not require sterilisation and hence is safe for a child to consume and reduces their exposure to contamination risk from environmental contaminants such as dirty water.

The IHS2 includes data on the current age of mothers but crucially unlike the DHS there is not data available on the spacing of births or breastfeeding rates and as such these factors cannot be considered; although a proxy for birth spacing is included which is outlined in *Chapter 4*.

Utilisation of health services

The importance of health services in influencing health outcomes cannot be ignored. Caldwell (1979) indicates that public health services *"undoubtedly reduce mortality"*[28] and they are a key component of the Mosley and Chen (1984) framework. Health services range from specific care associated with pregnancy including antenatal, delivery and postnatal care, as well as, immunisations and other direct child health interventions such as nutrition programmes. It is clear why health services are important in determining health outcomes; prevention, detection and cure. As our knowledge of

[28] p.396.

disease and illness has grown, so too has our understanding of how we can prevent the spread of disease. This can be in the form of very simple steps such as washing our hands and covering our nose when we sneeze. We know how certain diseases are spread; HIV for example is spread via the exchange of bodily fluids. If we know how disease spreads, we can seek to take steps to prevent diseases spreading in the first instance.

We may also undertake direct intervention, in the form of immunisation against certain conditions. Immunisation is a process which seeks to make an individual "*immune or resistant to an infectious disease"*[29] by causing our own immune system to develop antibodies against the condition being vaccinated against. The WHO indicates[30] that immunisation is one of the most cost-effective investments in health that we can make and can be delivered in a variety of environments on a large scale. The Extended Programme on Immunization (EPI) was established in 1974 in order to continue the expansion of vaccination programmes across the world. The success of immunisation should not be underestimated; it led to the eradication of smallpox and substantial gains towards the eradication of polio with only 716 confirmed cases reported in 2011 compared to 350,000 cases in 1988. [31]

Immunisation is believed to prevent between 2 and 3 million deaths every year from conditions such as diphtheria, tetanus, pertussis[32] and measles. To put this into context, the WHO suggests that about 29% of all childhood deaths are vaccine preventable.[33] The number of countries with 80 percent or more of the population receiving vaccinations against diphtheria-tetanus-pertussis (DTP3) is now 162 countries worldwide. In addition a number of countries are now seeking to use a new range of vaccinations; including vaccinations against Hepatitis B, rubella, mumps, yellow fever and maternal and neonatal tetanus (MNT).[34]

[29] http://www.who.int/topics/immunization/en/
[30] http://www.who.int/topics/immunization/en/
[31] http://www.who.int/immunization_monitoring/Global_Immunization_Data.pdf
[32] Whooping cough.
[33] http://www.who.int/immunization_monitoring/Global_Immunization_Data.pdf
[34] http://www.who.int/immunization_monitoring/Global_Immunization_Data.pdf

Where women have their children is also an important factor in determining child deaths, particularly those deaths which occur soon after delivery. A large proportion of deaths occur very shortly after birth; around 40% of deaths occur within the first 28 days.[35] Ensuring women have access to healthcare and appropriate medical care is therefore important. Lavy et al. (1996) find that there is a negative relationship between access to child health services and child mortality. Moreover, Feng et al. (2011) find that the probability of child mortality is lower when the baby is delivered in a hospital compared to at home. Therefore, it is important to take into consideration the health services that women may have utilised during pregnancy and at birth in order to evaluate the relative importance of these factors. One important issue to note is that there may be a huge variation in the quality of the healthcare services that women access and this is something that is not picked up on within estimation results. I cannot infer whether the care received at one hospital or in a clinic is better in one community relative to another. Therefore, the estimation results that are presented essentially assume uniformity in the provision of these health services which is unlikely to reflect reality.

Ideally one would wish to be able to control for such differences in service provision; it may not be the service per se that is influencing the outcome but the quality of that service. This is beyond the scope of this thesis but it is important to acknowledge that a hospital in Malawi is very different to a hospital in the UK and may very well be different to the provision in other areas of sub-Saharan Africa. Therefore as Huynen et al. (2005) suggest, contextual awareness is relevant when considering the potential value of different interventions.

The provision of suitable antenatal care seeks to ensure that good hygiene practices are maintained during delivery and after birth to prevent the spread of infections. Midwives and other medical staff can also observe a mother and her child to detect whether there are any concerns which need to be addressed. If action is then required it is essential that facilities have the right equipment and medication to provide treatment. In cases where ill-health cannot be prevented, the next most important step is to treat it. The aim

[35] http://www.who.int/pmnch/topics/part_publications/201112_essential_interventions/en/index.html

in such a case is to ensure that morbidity is minimised and mortality is avoided wherever possible.

Location can also play an important role in explaining child mortality rates. Notably, rates are commonly found to be higher in rural areas within developing countries (Wang, 2003). Pandey (1998) proposes that health services are more difficult to access and are of a poorer quality in rural areas. Therefore, health outcomes may be less favourable for rural residents. Pandey (1998) finds evidence that child and infant mortality is higher in the rural area and this effect can be quite large in some cases. However, once additional controls are added, this effect disappears. This suggests that differences in outcomes between the rural and urban area can be explained by factors which are related to residence, but not particularly residence by itself. This finding is similar to that of others who find that child mortality is higher in rural areas but that there are underlying factors which lead to this result. For example, Rahman and Sarkar (2009) find that child mortality is higher in rural areas within Bangladesh. This is attributed to higher levels of fertility, lack of access to health services and poverty within the rural area. Sastry (1997) also finds that child mortality rates vary by area of residence, but this variation can be explained by differences in underlying socioeconomic characteristics, particularly levels of maternal education.

Controls for the area of residence are therefore included within the subsequent analysis in a number of different formats. Given that the IHS2 data identifies location not just by a rural and urban split but also by region and a larger subset of locations, the analysis that follows in *Chapter* 4 considers all such variations in order to explore differences in child mortality within Malawi which may be sensitive to the location indicator selected.

2.3.3 Conclusion

As I indicated at the beginning of this section there are a number of determinants of child mortality/survival which are considered within the literature. It is clear that there are some areas of agreement and some areas of disparity. One area on which there seems to be a difference of opinion is in regard to the relationship between income and child health. I anticipated that the literature would suggest that improved income would lead to better health outcomes. Whilst there is some evidence to support this, it is certainly not a forgone conclusion from what I have observed during this review. However, as it is unclear exactly whether there is a relationship or not, I will continue to consider it as an independent variable within the empirical work which follows.

One of the problems in using income data in a developing country context is that it is often self-reported and not verified by formal records in household surveys or it is national data which prohibits analysis as the micro level. Self-reported income data could therefore suffer from considerable reporting error, especially if the recall period is lengthy and if education levels are relatively low as people may not keep accurate records. What is apparent in the IHS2 is that income from different sources only refers to recent periods in time; the last month or the last cropping season. The result of this is that the data could be time sensitive and may not be representative of the households' income more generally.

There are different views on the pathway or mechanism which maternal education operates through, in terms of reducing child mortality rates. However, whilst I do not dispute that improving our understanding of how maternal education affects child health outcomes, it is not the focus of this thesis, but it is certainly a topic for further research. Moreover, whilst there is evidence to suggest that location is a determinant of child mortality which cannot be ignored, there are several pieces of research which suggest that it is the proximate factors, which are operating via location, in determining child mortality rates.

Our ability to examine the determinants of any explanatory variable is often very much down to data availability. In *Chapter 4*, I note where there are some variables which, based upon the literature, I would like to have included in my analysis but have been unable to do so, due to data restrictions.

Based on the discussion within this section, the empirical work I conduct in *Chapter 4* is based upon the framework developed by Mosley and Chen (1984) examining the impact of some of the proximate determinants upon child and infant mortality rates in Malawi.

2.4 Wages and health

This next part of the discussion moves on to the human capital theory literature which is the basis of the empirical work undertaken in *Chapter 5.* The human capital literature developed as it became increasingly noticeable that gains in income were being generated that were not reflected by increases in physical capital (Becker, 1962; Shultz, 1961). The main areas of human capital identified by Shultz (1961), Becker (1962) and Ben-Porath (1967) are; health, on-the-job training, formal education, study programmes and migration. The focus of the *Chapter 5* is an examination of the relationship between health as a form of human capital and the market outcome; the wage rate in Malawi. The initial discussion presented will consider human capital in a broader manner to provide an overview of the human capital literature as the model utilised within the chapter is based on a Mincer (1974) earnings equation.

If investing in this form of human capital was completely costless, then Becker (1962) argues that there would be an unlimited demand for such investments. The truth of the matter is that investment in any activity which increases human capital stock is not costless; it requires the diversion of time and resources from other activities. The rate of return from the different forms of investment does, however, differ, although all make positive contributions in raising the income potential of individuals (Becker, 1962).

Becker (1962) suggested that many people improve their stock of human capital whilst in employment; we learn new skills and indeed polish the older ones. Investing time in learning a new skill raises the productivity of the individual over time; they are more able to complete a task in shorter periods of time and undertake a greater range of tasks. In such a case the marginal product of the worker and the wage are equalised in each time period. The introduction of training changes this one period relationship; current output may be lower but the anticipated gain is increased output in future following the completion of training. The type of training that people may undertake can be general, where skills are transferable to other industries/activities; or specific, where the predominant gains to such training are maintained within the firm according to Becker (1962). The potential disadvantage to firms of general training is that, such skills may not remain within the organisation and therefore the firm may not benefit from its investment. However, Becker (1962) pointed out that firms can circumvent this difficulty by moving the burden of the cost from themselves to the employee; lowering wage offers in the current period to cover the cost of investment. The cost of the investment is then shifted to the employee; the opportunity cost of them training is now reflected by accepting a lower wage offer (Becker, 1962).

Human capital, can be treated like physical capital in the sense that it depreciates over time; at a young age people accept a lower wage (whilst training) in exchange for higher wages in the future (once the gains in marginal productivity can be realised) but the rate of change of the gain decreases over time and is often captured within Mincer (1974) type earnings equations by the inclusion of experience. Specific training differs in that a higher proportion of the training will be associated with the activities of the firm; skills in this sense are less transferable to other firms (Becker, 1962). Given that such training provides fewer transferable skills, Becker (1962) observes that wages tend to be higher (relative to those receiving general training) which reduces the incentive for people to leave and also reduces the incentive for firms to dismiss such workers.

Within the estimations presented within *Chapter 5*, it is not possible to include an indicator for training but data is available on the age of individuals which can be used to create a proxy measure for experience which may capture on-the-job gains in human capital. Using a broader measure such as experience as an indicator for on-the-job training is not ideal as it does not capture the variation in training that individuals may have received over their working life but it is a commonly used indicator within earnings estimations and therefore has comparability advantages and indeed such information can be extracted from the IHS2 dataset.

A similar and commonly considered form of human capital explored with the literature is the role of education and Becker (1962) describes this as an organisation that specialises in producing training. Training in this sense can again be specialised for particular fields of work or be general. Education costs are associated with factors such as fees and equipment costs which are deducted from earnings. Indirect costs are again associated with forgone earnings; one cannot be in two places at once and the effect of education is very much like that just presented for on-the-job training; education raises the marginal productivity of people and translates to higher wages. There are some options to consider when examining the amount of education that an individual may have; years of schooling or the level of education achieved which is measured by qualifications. Shultz (1961) argues that years of schooling are not the most ideal measure to use as there has been a shift towards higher levels of education. Therefore comparing the number of years across individuals is unlikely to compare like with like. Consequently, one may consider levels of education as an alternative which is the approach undertaken within the empirical chapters. The number of years of schooling an individual has is reliant on accurate information on the age at which someone first attended school and when they left; reporting error for either of these dates would lead to error in the reported number of years of schooling. In comparison, individuals could report the highest level of education that they completed. However, to explore the sensitivity of the results to the measure of education selected, one can consider both variations of education in the subsequent analysis.

The final topic considered here is that of health as a form of human capital. As the Grossman (1972) model illustrates investment in health which is a component of human capital has both consumption and investment attributes and this is also true of the other forms of human capital according to Shultz (1961). Investment in health capital increases the number of healthy days available for work and the efficiency of workers as they may be stronger (Becker, 1962). As the discussion of the Grossman (1972) model indicates, investment in health is in the form of health services which includes things such as medical care and better nutrition. Firms as well as individuals can invest in health services, by providing staff with breaks, medical examinations and safety equipment such as goggles and gloves for hazardous jobs (Becker, 1962).

Or put more simply, there are quantitative and qualitative gains from increasing the stock of human capital (Schultz, 1961). The accumulation of human capital is therefore a reflection of the skills and knowledge that people possess, which enables them to be more productive which is valuable in the marketplace (Shultz, 1961; Becker, 1962). Such investment is not costless, however; if people are to undertake investments, then there must be some form of compensation for the time, effort and resources that are diverted away from other activities in order to accumulate capital.

Thomas and Strauss (1997) argue that whilst there has been a great deal of work on the effects of education on wage outcomes, there are other important human capital factors which have not been the focus of research, particularly health within developing countries. One reason they give for this, is that often, there is insufficient empirical evidence to support the view that there is a relationship between health and productivity. The reason for this is that, surveys often don't include measures of wages and/or productivity which is also noted by Ghatak (2010). This is where the exploration of the IHS2 for Malawi presents a potential opportunity; wage data is available and so too is information on the ganyu labour wage which is a form of casual income in the rural setting and presents an additional line of consideration relative to the existing literature which has primarily considered urban workers. However as previously noted, this data may be subject to reporting error in the absence of formal records. The recall period related to wages in the IHS2 refers to the most recent payment which therefore seeks to avoid a lengthy recall period. A potential problem with this approach is that the wage reported may not be reflective of the wage someone may receive over time that it more “typical”.

In practice, one of the real challenges to examining the relationship between health and economic outcomes is that of the causality of the relationship even without the additional complications of suitable data. Thomas and Strauss (1997) highlight that *"any component of income, such as wages or labour supply, may affect current behaviour which, in turn, affects health….and vice versa".*[36]

As this field develops, Thomas and Strauss (1997) indicate that of those studies that have been conducted, many consider particular sub-groups within a country such as only male rural workers and use a limited number of health measures (Ghatak, 2010). This presents limitations as the results may be sensitive to the choice of health measure and only relevant for particular groups. In order to move forward in our knowledge and understanding, we need to ensure that surveys ask the right questions to elicit the information that we need to explore this area of research. In particular, to develop cross country comparisons we need to consider consistent and comparable measures of health status, reliable income data and the inclusion of different types of groups rather than just

[36] p.160.

urban male workers. Moreover, given the causality and endogeneity issues between health and income, we need to consider what other data we might need for econometric purposes so that we may construct suitable indicators for IV (instrumental variable) estimation, for example.

2.4.1 Examining health and productivity outcomes in practice

As the discussion above highlights, people make investments in themselves, in different forms of human capital such as health, education and training The question is what is the value of these investments to individuals and what is the "cost" in terms of forgone wages to the individual as a result of a loss of health human capital in particular.

Willis (1986) indicates that earnings functions are used to estimate the effect of different attributes on wages and can be applied to a whole range of different issues; which can include gender or race discrimination. The regression coefficients provide an indication of the rate of return to that particular investment. The standard form of an earnings equation developed by Mincer (1974) considers the logarithm of wages as a function of characteristics such as education and experience plus other observable covariates.

Thus in equation (1), s represents the years of schooling undertaken, x represents post schooling years of experience and the square of experience captures initially increasing and then diminishing returns to experience (assuming here that we shall find $\beta_3 < 0$).

$$In\, y = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 x^2 + u \qquad (1)$$

In addition to education and experience, the Grossman (1972) model highlights that we undertake investments in health as another form of human capital to generate additional healthy days and these days have an economic value; more time for work and leisure activities and working time is rewarded in terms of remuneration. Empirical estimates hence seek to estimate the value of healthy working time (or the lack thereof).

An important contribution to the health and productivity literature made by Thomas and Strauss (1997) is the use of a range of different indicators of health when they examine wage outcomes in urban Brazil. This enables them to explore the sensitivity of results given the measure of health used. Thomas and Strauss (1997) use a household budget survey which collected information between August 1974 and August 1975 to investigate the impact of four health indicators (per capita calorie and protein intakes, height and body mass index) upon wages, for urban workers.

In contrast to nutritional measures and anthropometric measures, I consider functional limitation as an indicator of health using the IHS2. The IHS2 does not include anthropometric measures of adult health. They argue that nutrient intake reflects current inputs into production, which is indeed recognised by the Grossman (1972) model, but this measure has substantial data requirements; which are just not available in the IHS2. Therefore a similar approach in terms of the choice of health measure is not adopted for these reasons. In particular for Thomas and Strauss (1997) to collect the necessary data for the nutritional measures, expenditure was collected on all food items which were then weighed and wastage calculated. Over 300 foods were included, followed by the calculation of nutritional values and then standardised composition tables. The final step taken was to convert these values into the number of people attending each meal.

The Brazilian dataset collected information on income, expenditure and socio-demographic indicators from 53,000 households and all of their members. Each respondent aged 14 years or over, reported their labour supply and their income for the previous year. Longitudinal data such as this is would be preferable to account for seasonal changes in income, but this is not available for the IHS2 and therefore in contrast I consider the most recent wage data. In addition, the Brazilian respondents also provided information on the number of hours they worked in the previous month and whether they had been employed in that job throughout the whole of the previous year. Similar data is reported for the IHS2 which then enables a daily wage rate to be calculated for the Malawian estimations. Based on the information provided about their employment over the last year, an average number of hours worked per month is used in the Thomas and Strauss (1997) analysis.

Thomas and Strauss (1997) separate waged workers from self-employed individuals as they believe that returns to health, may differ across different types of employment sector. In particular they note that, self-employed income is very difficult to measure, as it is often difficult to separate an individual from their business activities. Moreover, in family run businesses, it is also difficult to attribute income to one particular individual, if more than one family member works at the organisation. This problem can therefore be a significant impediment to assessing self-employment income in the rural sector where family owned enterprises are often farm based (Thomas and Strauss, 1997). Therefore, Thomas and Strauss (1997), specifically consider only the urban sector which comprises of just over sixteen thousand men and almost eighteen thousand women in the South and Northeast of Brazil. I consider all individuals reporting a wage but also separately examine ganyu labour which is an important source of casual labour income with Malawi to explore whether there are differences between the two groups.

Shultz and Tansel (1997) use a different approach, which is self-reported morbidity which Thomas and Strauss (1997) seek to avoid as they suggest that on the whole, morbidity has quite a small effect on income and wages, primarily because morbidity is often short term. Moreover, Thomas and Strauss (1997) note that self-reported morbidity is problematic in that there can be random and systematic reporting error. The advantage of the method is, however, that the information can be obtained relatively easily.

Schultz and Tansel (1997) seek to estimate the effect of adult morbidity on labour productivity in two West African countries; Ghana and Cote d'Ivoire which are both sub-Saharan African countries like Malawi. They use a Living Standards Measurement Survey (LSMS) and specifically use information from the health and labour use modules. This presents some comparability with the IHS2 dataset as this is also taken from the LSMS series and consequently has similar questions and enables me to present a similar approach to Shultz and Tansel (1997) for comparability. The information obtained from the survey enables Shultz and Tansel (1997) to examine specific short term periods of illness, the labour activity and wages of these individuals. Shultz and Tansel (1997) suggest that the self-employed are more likely to subjectively estimate the impact of illness on their working patterns as they can make adjustments for poor health more readily than a waged worker. Therefore, they suggest that estimating the impact of morbidity upon waged workers will be more accurate, as the incentive to continue working is stronger than for the self-employed. Moreover, they also reiterate the difficulties that can be encountered when trying to measure the earnings of self-employed workers and how particularly challenging this is in family based operations. Given that both Thomas and Strauss (1997) and Shultz and Tansel (1997) suggest that waged workers are a more appropriate group to consider, I too also only include waged workers in my analysis. An additional line of development, however, as previously mentioned is the inclusion of a separate group; ganyu workers.

2.4.2 Modelling approaches

Both Thomas and Strauss (1997) and Shultz and Tansel (1997) highlight the practical issues associated with examining health outcomes within the labour market; namely how one measures health and the problems that can be encountered when trying to capture this within a survey; as was considered earlier on within this chapter. Thomas and Strauss (1997) and Shultz and Tansel (1997) clearly use very different measures of health; with the work presented in *Chapter 5* leaning towards the latter which in some sense is due to data availability but predominantly because the self-reported measure used by Shultz and Tansel (1997) captures general morbidity and lack of ability to continue normal activities. This measure isn't associated with very particular activities such as stair climbing or mopping floors; instead people can report whether or not they could work and if not, for how long which provides an indication of the severity of morbidity. Human capital theory argues that investment in ourselves has both quantity and quality effects; Shultz and Tansel's choice (and also my own choice) of health (specifically a lack thereof) seeks to capture these effects to a degree, without the need to use a more sophisticated method such as a DALY.

An issue noted in both papers is the direction of causality between the selected measure/s of health and the labour market outcome which could be wages and/or productivity. The premise put forward in both papers is that better health can lead to a worker being more productive and this in turn may translate into higher wages/income. Those individuals with higher income then have the funds to make investments in their health and so this becomes a reinforcing mechanism. Both Thomas and Strauss (1997) and Shultz and Tansel (1997) suggest that ordinary least squares (OLS) may not be an appropriate method if the outcome of interest (wages, productivity and health) is endogenously determined. Thomas and Strauss (1997) use IV estimation whereas Shultz and Tansel (1997) use both OLS and IV estimation which is the approach I follow in *Chapter 5.*

Specifically, Thomas and Strauss (1997) examine the effects of health upon productivity and wages using four different measures of health: BMI, nutritional intake, height and weight, starting with the use of a reduced form labour supply function. They also include a vector of observed exogenous individual characteristics which is a standard Mincer (1974) type approach that I also follow in *Chapter 5*, plus the addition of a vector of community-level characteristics including prices, local demand and infrastructure.

As they suspect that there may be correlation between the health variables and the error term, they use an instrumental variable (IV) approach. Suitable instruments are those which affect health but not wages. This includes measures at the community level (in their model) such as the disease environment, health infrastructure and the price of health inputs. The argument is that these indicators are related to health outcomes but do not directly influence wages. Following the presentation of the OLS results and a test for endogeneity, I also consider IV estimation and include indicators such as those presented by Thomas and Strauss (1997) and Shultz and Tansel (1997).

As previously indicated, Shultz and Tansel (1997) adopt a similar type of approach except that instead of the variety of measures used by Thomas and Strauss (1997) they use a measure known as disabled days as their indicator of health status. This is also the approach taken in *Chapter 5*. In the case of the Shultz and Tansel (1997) study, respondents were asked whether they have experienced any ill health in the last four weeks and if they respond with a *"yes"* they are then asked how many days they were unable to undertake their normal activities for. Therefore, if someone is ill and unable to work in the reference period, for each day that this is true, it is classed as a disabled day.

In the Shultz and Tansel (1997) model, it is assumed that health is an accumulation of investments over time; this includes investments made in childhood such as nutrition and vaccinations and also includes current health care to deal with chronic and/or acute health conditions. An individual's wage is determined by health, other commonly associated endowments such as education and other controls including experience which is again a standard Mincer (1974) approach. In a similar vein to Thomas and Strauss (1997) they consider food prices and local health infrastructure as instruments. In addition, Shultz and Tansel (1997) also include some additional control variables such as distance to the market, rainfall and interactions between the time of year when the survey took place and the local severity of malaria related health problems. Such considerations are not included within *Chapter 5* due to lack of data availability.

2.4.3 Empirical results

The LSMS data used by Shultz and Tansel (1997) was collected in the 1980s during a series of waves; where around half the original respondents were subsequently interviewed at a later date. Approximately one fifth of all men and women aged 15-39 reported that they had been ill or injured in the last 4 weeks in Côte d'Ivoire and this rises quite sharply for those aged over 39 years of age; almost two-fifths report a bout of illness or injury (Shultz and Tansel, 1997). The average number of days an individual is ill for also rises with age for both males and females; from 1.04 for rural males aged 15-39 to 4.44 for those aged 40 or more.[37] This pattern is similar for females; an average of 1.37 days of illness are reported for rural females aged 15-39 and 3.83 days on average are reported for rural women aged 40 or over within a four week period.[38] The average number of disabled days is notably higher amongst those aged 40 or over regardless of where an individual lives or their gender.[39]

A similar approach was adopted in Ghana using the LSMS. One noticeable difference is that the proportion of people experiencing illness or injury is much higher in Ghana than in Côte d'Ivoire across all age groups (Shultz and Tansel, 1997). The average

[37] Not separated by education level. Results in the average number of disabled days differ across levels of education but the basic pattern between age groups is very similar.
[38] As per previous footnote; differences in education have not been accounted for.
[39] As per previous footnote; differences in education have not been accounted for.

period of illness reported by men and women, does not differ substantially on the whole, between Ghana and Côte d'Ivoire (Shultz and Tansel, 1997). However, differences between the numbers of disabled days reported are more pronounced when looking at differences between the genders for those aged 15-39. It is reported that women aged under 40 years of age, in both countries, experience more disabled days than a comparable male. This is not unexpected as of course, women aged less than 40 years of age could very well be childbearing. Therefore at least some of the difference, could be attributed to pregnancy related health issues, rather than health problems more generally. The results presented in *Chapter 5,* present comparable tables so that similarities and differences may be noted, particularly given that the approach is very similar and considers another sub-Saharan African country.

In addition to the differences between morbidity, Shultz and Tansel (1997) also consider differences in mortality rates across different groups. They find that those living in urban areas report lower levels of mortality, as do more educated individuals. These findings are consistent with the theory that those who are more educated may be more able to recognise ill health in themselves and their children. Therefore such individuals would seek medical advice more readily than someone who has less education. Moreover, individuals with a higher level of education may also be in better paid jobs which raises the return to healthy days. Therefore, the cost of being ill is greater which the Grossman (1972) model predicts. This could motivate an individual to seek medical attention sooner as well. In addition, those living in the urban area may be able to access medical care more readily than an individual in the rural area; rural inhabitants could experience greater challenges to accessing medical care, which could be physical and/or financial.

How people perceive their health in terms of what is "normal" for them leads to a great deal of variation in the reporting of morbidity, when using a self-reported measure as previously noted. Shultz and Tansel (1997) find no evidence to support the premise that morbidity rates are higher amongst what they refer to as the *"elite"*[40] which stems from the *"cultural conditioning"*[41] hypothesis of Johansson (1991). Those who have the

[40] p.268.
[41] p.268.

greatest to lose may be more sensitive to reporting morbidity. Equally, these individuals could also seek to undertake additional preventative measures, to ensure good health. Therefore such individuals report less morbidity.

Shultz and Tansel (1997) limit their results to only men as only a small proportion of female survey respondents in both countries are employed in the wage sector. Therefore, female respondents were not considered to be a representative sample. Thomas and Strauss (1997) report their findings for males and females separately as a much larger proportion of Brazilian females work in the labour market and therefore it is possible to examine both groups.[42] The results presented in *Chapter 5,* consider both male and female wage workers separately and I too also find that the female sample is much smaller than the male sample.

Shultz and Tansel (1997) initially present OLS results which examine the determinants of disabled days in the two countries. The sample sizes for Côte d'Ivoire and Ghana are 1047 and 1460 respectively. The coefficients on education are as one would expect for primary and post middle levels of education; indicating a reduction in morbidity. However, none of the coefficients on education level are statistically significant at the 5% significance level. The results on different measures of household wealth, food prices and community health problems are also very mixed, with no clear pattern emerging, in terms of explaining the variation in the number of disabled days that individuals experience. One particularly interesting result is that the coefficient on the distance someone has to travel to a hospital is negative. Therefore, this indicates that the further away from a hospital someone is, the fewer disabled days they will have. However, the coefficient on the distance one travels to a doctor or nurse is positive. This rather unusual result is found for both Côte d'Ivoire and Ghana.

The IV estimates suggest that wages are 10.5% lower for each additional disabled day an individual experiences in the case of Côte d'Ivoire and 11.7% lower in Ghana which is a surprisingly large effect for one day of illness unless one assumes that disabled days

[42] 60% in the sample. Of these two-thirds work in the market sector and one third are self- employed (Thomas and Strauss, 1997).

are able to act as a proxy for general morbidity and therefore reflect a reduction in health capital.

The productivity results are then presented by Shultz and Tansel (1997), who examine wages, hours and earnings. The first panel of their table using maximum likelihood estimates and the actual reported number of disabled days shows that an additional disabled day has a negative effect on all three productivity measures in Côte d'Ivoire. Moreover, additional disabled days also reduce the number of hours individuals work by a small amount; although Shultz and Tansel (1997) note that this is not estimated precisely. In their final table, Shultz and Tansel (1997) consider the effect of disabled days squared rather than a linear specification of disabled days. They find using this specification and an IV estimate that an additional disabled day in Côte d'Ivoire leads to a 33 percent reduction in the wage received. They suggest that this is the equivalent of a 65% reduction in annual earnings given a 32% fall in hours. The results for Ghana are also very startling; an additional disabled day leads to a 26% fall in the wage an individual receives and this translates to a 32% reduction in earnings given a 21% fall in hours (Shultz and Tansel, 1997).

Thomas and Strauss (1997) examine the determinants of wage rates (per hour) in the urban wage market in Brazil using IV estimation in their first set of results. Notably, their first model only considers the effect education has on wages. Possessing additional education; measured in terms of an additional level of education relative to someone having no education at all, is shown to have a large and positive effect on the wage rate received. Thomas and Strauss (1997) then go on to add some additional indicators for health. They find that being taller and heavier have a positive effect on an individual's wage. Height reflects an accumulation of previous health investments, whereas BMI (due to weight) is partially determined by current behaviour. The results suggest that people who are healthier (taller) and heavier (greater BMI) receive a greater wage (Thomas and Strauss, 1997).

Two additional measures are also used; calorie and protein intake. An increase in both calories and protein intake is associated with a positive effect on wages. However the results suggest that the effect of an increased calorie intake is more substantial when the initial calorie intake is at a very low level. The final model from their first set of results presents all the different indicators of health status within one model as each indicator captures slightly different aspects of health and concludes that they all positively affect male wage rates in Brazil. Thomas and Strauss (1997) also examine those who are self-employed and also include female workers. The results vary slightly and are less clear for self-employed individuals, however; they conclude that in general overall good health generates a large return on wages.

In both papers, selection bias is considered as a potential problem. The reason for this is that in both cases particular groups are examined.[43] Therefore it could be the case that there are certain characteristics that determine whether someone enters into the wage market and these unobserved or unaccounted for factors affect the estimates reported. Shultz and Tansel (1997) do not find evidence to support that sample selection is a substantial problem in Côte d'Ivoire but they do find evidence that it is problematic in Ghana. Therefore, in order to be able to compare their regression estimates between the countries they only consider the maximum likelihood estimates which are corrected for sample selection. Thomas and Strauss (1997) also test their results to see whether the decision to enter into the wage sector affects their regression estimates. They conclude that ignoring selection into the wage sector does not affect their results.

In addition, both papers report on the problems that they have encountered in trying to identify suitable instruments given the potential endogeneity of the health measures they use. Thomas and Strauss (1997) indicate that caution should therefore be used when interpreting results. Shultz and Tansel (1997) conclude that longitudinal data would be beneficial as this would enable researchers to examine changes in individuals and their environment and therefore develop instruments reflecting these changes over time for the individual. Shultz and Tansel (1997) also note that even when surveys are repeated in waves; very often the individuals are not matched correctly over time even for basic information such as gender and age, which creates an impediment to progress.

[43] Workers rather than the entire population.

2.4.4 Conclusion

There are several problems which researchers encounter when conducting empirical work in the health and productivity literature. Problems that can be encountered include; identifying the direction of causality, data availability with regard to indicators of health and how to deal with the endogeneity of health in terms of identifying appropriate instruments. Our selection of a measure of health may often be determined by the data which is available which can limit the breadth of our definition for health. The choice of health measure therefore determines a particular dimension of health and as we have seen; different measures of health measure different dimensions of health. Excluding certain aspects of health therefore makes any results from empirical work highly sensitive to the choice of health measure chosen. The measurement of wages or different sources of income is also problematic due to a lack of formal records, potential reporting error and relatively short reference periods. Future research within this field therefore needs to consider the surveys that are being used. If we don't ask a variety of questions about health then we are limited in what we can say about the effects of good or bad health on a whole range of issues.

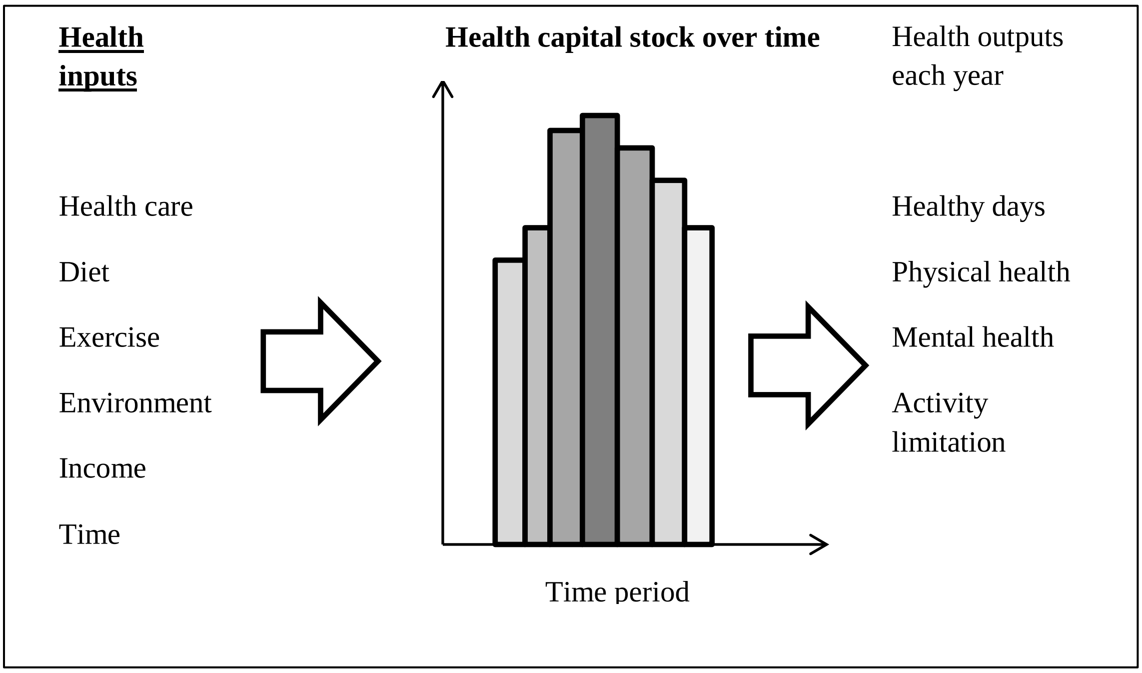
2.5 Health shocks and the household

Gustafsson-Wright et al. (2011) highlight that; health shocks have *"devastating and long-lasting consequences, especially for poor households"*.[44] Moreover, Wagstaff (2007) highlights that in the World Bank publication *Voices of the Poor* there is a large amount of anecdotal evidence regarding the devastating effects that health shocks have on households. *Chapter 6* therefore seeks to utilise the Grossman (1972) model which has been previously outlined to investigate whether health shocks do have a realisable effect on household incomes in the KwaZulu-Natal Province of South Africa. Figure 2.5.1 summarises the Grossman (1972) model. Using the Grossman (1972) model, the proposition is that if there is a decline in health capital, then the outputs will also be less favourable in the next period ceteris paribus. The consideration of a health shock in the context of a household unit is different to the use of morbidity at the individual level which is the focus of *Chapter 5,* as Lucas and Bloom (2006) argue that the use of self-

[44] p.143.

reported morbidity for relatively short periods of time fails to address the main issue which they believe should be of primary concern: "*the medium and long term impacts of health shocks and the impact of illness on sustaining normal household functioning".*[45]

Figure 2.5.1Illustration of the Grossman (1972) model of health



Source: Reproduced from Pruckner (2010)[46]

Chapter 6 therefore seeks to explore whether there is evidence that health shocks have a longer term impact on households, specifically in terms of income, which is used to purchase health inputs in the next period. The next part of the discussion will outline what health shocks are, including different indicators of health shocks from the existing body of empirical literature. I will then go on to consider the specific economic outcome of interest within this thesis which is income.

2.5.1 Defining health shocks

There is a range of different views on what constitutes a health shock and an important point to raise at this juncture is the fact that many refer to an individual. The KIDS data

[45] p.5.
[46] http://www.econ.jku.at/members/pruckner/files/lehre/health/demand.pdf

utilised in *Chapter 6* considers shocks at the household level which is explained in more detail within the chapter.

Riphahn (1999) defines a health shock as a sudden deterioration of health brought about by an accident or disease and uses a self-reported health satisfaction ranking system, whereby someone rates their health on a scale of one to ten. A fall of five points on this scale is then classified as being representative of some form of health shock. The KIDS dataset does include a question on changes in health over a 5 and 6 year period respectively for each core[47] household member. The scale used within KIDS is a five point scale which gives a narrower range for determining a health shock; instead of five points, I consider a decline of two points on the scale.

As an alternative, Wagstaff (2007) uses the death of a working-age head, a prolonged period of hospitalisation and a large drop in BMI of the household head as indicators of health shocks. Deaths of household members are considered within *Chapter 6* in line with the approach of Wagstaff (2007). Crucially, however, the specific individual is not identifiable; the death of working/productive members of the household is highly likely to influence the effect upon the household and therefore controlling for this would be preferable.

Garcia-Gomez et al. (2010) define a health shock as an acute hospitalisation which is not scheduled, cannot be postponed and immediate treatment is necessary. The KIDS data does not include any information about hospitalisations and therefore such an approach cannot be considered in the analysis that follows. Gustafsson-Wright et al. (2011) also consider hospitalisation as an indicator of a health shock alongside some self-reported indicators such as weight loss. Gan et al. (2006) define a health shock as one which requires a large amount of expenditure[48] or a hospitalisation.

Ideally, Riphahn (1999) suggests that information on the suddenness, severity and duration of the shock are desirable. The reason for this is that we can then separate out minor health ailments which fundamentally are not a priority concern from severe

[47] For reference, this is defined in Chapter 3.
[48] 5,000 RMB.

conditions or situations referred to by Lucas and Bloom (2006) as *"catastrophic illness"*.[49] This term sounds very simple in principle and suggests that the health issue is serious and therefore excludes minor conditions so in many respects "fits" the ideal criteria put forward by Riphahn (1999). The problem is that different groups such as the medical profession, insurers and the individual use this phrase in different ways (Lucas and Bloom, 2006). Therefore this term is very subjective and not as clear cut as it may first appear. The KIDS dataset includes a specific module on health shocks and the question within this module is worded to indicate that the event was a surprise. The question on differing shocks also seeks to gather information on the duration of the shock which is subsequently explored to observe whether the duration of a shock matters.

Lucas and Bloom (2006) highlight that many of the definitions of health shocks, may not take account of recurrent acute and/or chronic health conditions which of course would erode the stock of health capital over time. However, they suggest that these health conditions may be equally as problematic for the household even if they are not categorised as a "shock". Just because the health condition does not fit into a definition of a health shock, it does not mean that the impact on the household is not large. Therefore we need to be cautious or at least aware that certain definitions may lead to the exclusion of some conditions and underestimating the burden of morbidity for such households could be highly misleading (Lucas and Bloom, 2006).

2.5.2 The impact of health shocks on economic outcomes

Wagstaff (2007) identifies a number of factors which need to be considered when examining the effect of a health shock on income. The first issue he raises is that household earned and unearned income may behave differently in the presence of a health shock. Earned income will fall as the individual experiencing the health shock has fewer healthy days available; they may be completely unable to work or limit the amount of work that they do in order to aid recovery. At the household level, this may be compensated for by adjustments by others (Wagstaff, 2007; Gustafsson-Wright et al., 2011, Gan et al., 2006; Beegle, 2003). One response could be that other members of the

[49] p.1.

household work longer hours (assuming that there was not underemployment originally). Income data is reported by the household for the last month in KIDS from a series of different sources and is again not verified by formal records.

How severe losses are will depend on the age and characteristics of the individual experiencing the health shock. If the health shock is experienced by a child then earned income losses would be minimal/non-existent. The exception to this would be if the child was working and/or a working individual had to withdraw from the labour market for a prolonged period of time to provide care. Equally, if the individual experiencing a health shock was very old and/or not active in the labour market then one would also not expect to observe a fall in earned income. At the other end of the spectrum, for a working aged individual, the financial loss could be very large, particularly if others had to withdraw to provide care. In such a case it could be more than one income which is lost (Garcia-Gomez et al., 2010; Gan et al., 2006, Beegle, 2003).

If we think about the household as a collective unit, the differences in the losses households experience could also be explained by the type of jobs that household members are employed in. If someone received any kind of medical care or sickness benefit from their employer, even if only for a limited time, this would reduce any losses experienced by the household (Wagstaff, 2007). Such arrangements are far more likely to exist in urban areas and therefore, urban households may be more protected against income losses relative to rural households (Wagstaff, 2007). On the other hand, those residing in the rural area may be subject to more flexible working conditions which would enable people to work and care for a sick person at the same time (Wagstaff, 2007; Beegle, 2003).

The discussion presented suggests that some differences in the experiences of individuals or specifically households within the context of *Chapter 6* are dependent upon their characteristics. Therefore, controls for location, employment status and age are desirable and included within the analysis presented.

The behaviour of unearned income in the presence of health shocks is more difficult to quantify as it depends on a host of different security nets that may be in place; these could be formal or informal. In many countries formal protection may not be available even in urban areas (Wagstaff, 2007). Therefore it is more likely that the social security nets that catch the most vulnerable are informal. Riphahn (1999) includes an indicator for formal support in her study of the German labour market and finds that even with such support households are not fully protected against the loss of income following a health shock.

Support could come from a variety of sources such as friends, extended family members or from within the wider community. However, Lucas and Bloom (2006) note that there is not a great deal of empirical evidence examining these mechanisms and the effect that they have upon afflicted households. Putting this aside, support can come in many different forms, such as loans, gifts or in-kind support. For example, following a fall in earned income, unearned income in the form of remittances from other family members may rise to compensate for the initial losses (Wagstaff, 2007; Pityanon et al., 1997). It is also possible that the informal sector may be more willing to assist a household in times of hardship due to ill health relative to other types of shocks. For example Wagstaff (2007) suggests that friends and family members may be willing to assist a household which has come upon poor times due to a decline in health more readily than one which has made a poor decision leading to a loss of income and which they are accountable for. Moreover, formal credit may also be very difficult for households to obtain as traditional institutions would require some form of collateral or guarantee which many households may not have (Wagstaff, 2007). In the analysis presented in *Chapter 6*, unearned income is also considered and includes income from sources such as remittances and grants in order to observe whether there are changes in unearned income as a result of a household experiencing a health shock in the KwaZulu-Natal province.

In addition to income, health shocks can affect other economic outcomes which may not immediately be considered. Lucas and Bloom (2006) suggest that an alternative consideration is to examine the effects that health shocks have on expenditure. However, they note that in practice poor people have very low levels of expenditure on certain items, particularly healthcare, and therefore it may be very difficult to observe changes or differences in expenditure levels amongst certain groups (Lucas and Bloom, 2006). The KIDS dataset does include modules that collect information on expenditure, which is separated into food and non-food spending for the past month. This indicator is not however explored within the thesis as the focus is specifically on income.

Wagstaff (2007) proposes a different suggestion. He considers that an important variable to examine in response to health shocks is food consumption. This is because he argues that it is a commonly held belief that high medical costs can lead to a large reduction in consumption and particularly food consumption. Pityanon et al. (1997) find evidence that households with very limited resources such as savings or assets reduce their consumption. Therefore, a concern is that if the reduction in food consumption is significantly large then this can lead to people being too weak to work, due to malnourishment (Pityanon et al., 1997; Wagstaff, 2007). There is only one question related to food consumption within the KIDS dataset which only seeks to establish whether or not the food was consumed by the household. No further information is presented and therefore it is not possible to explore this in any additional detail.

2.5.3 Coping strategies and the effect of health shocks on economic outcomes

Households experiencing health shocks (or any other type of shock for that matter) will seek to make adjustments in order to minimise the disruption to the household. However, if health shocks lead to households employing coping strategies which have negative effects on other aspects of well-being then we need to consider all the costs and benefits which could be associated with any intervention.

A message from Lucas and Bloom (2006) is that whilst we can make investments to improve health, the question that should also be asked is whether we should, given that there are wider implications associated with interventions which may be overlooked. Their concern is that households could invest in treatment but in so doing, there could be knock on effects within the household. For example, current consumption could be forgone in order finance treatment costs. However, this could have potential repercussions for other household members both now and in the future.

Wagstaff (2007) indicates that households facing income shocks will respond by seeking to smooth their consumption. However, poor households are less able to smooth their consumption as they have limited access to savings, borrowing and insurance which are three common coping strategies employed by households. However, there is evidence that the poor are less well equipped cope with income shocks (Wagstaff, 2007; Gustafsson-Wright et al., 2011) and therefore this group is most at risk.

Insurance is one way in which people can protect themselves from an unexpected or undesired event. Gustafsson-Wright et al. (2011) examine the impact of health shocks on uninsured individuals in Namibia. The selection of Namibia is important as the authors indicate that the public healthcare system within the country is relatively well developed and therefore people can, in principle, access healthcare. Fees are charged at a flat rate depending on the type of clinic or facility that someone visits and drugs are heavily subsidised according to Gustafsson-Wright et al. (2011). However, whilst this should enable people to access care, it is also noted that the system is quite strained. Waiting times can be lengthy and highly qualified medical personnel are in short supply.

The survey data used by Gustafsson-Wright et al. (2011) includes information on socioeconomic characteristics as well as medical information. The socioeconomic information is gathered from a sample of the Greater Windhoek area covering 1,769 households and 7343 individuals. In addition to this, results from saliva tests to detect HIV for those aged 12 or more is used to assess health status. At the time of the study a number of households moved due to a new housing development in the area and therefore just over half of the original respondents who were tested could be traced at a later date.

Gustafsson-Wright et al. (2011) examine the take up of health insurance by different groups. A household is considered to be insured if at least one household member is subscribed to a medical aid programme. They find that only 5 percent of the poorest individuals have subscribed to health insurance whereas 70 percent of the richest are signed up to a healthcare programme. In addition, there are also differences in the take up of insurance depending on the type of job or employment sector that someone is in;

insurance is more commonly found for those who work in government or defence based occupations. What is very interesting is the fact that when the insured and uninsured were asked about their health status, the insured indicate that they experience more poor health than those who are uninsured. This result could be attributed to two factors (Gustafsson-Wright et al., 2011); the insured experience worse health and therefore seek to mitigate against losses by taking out insurance. Alternatively, it could be that the insured are better at identifying illness and/or they are more sensitive to their health status and therefore report a greater incidence of health shocks.

The findings from Gustafsson-Wright et al. (2011) show that insured individuals will seek to use healthcare provisions more readily than those who are uninsured. There are also gender differences; individual women are less likely to use health care services than men but a female headed-household is more likely to access health care than a male headed-household. Moreover, they find evidence to support the view that the education level of the household is important; a higher level of education is associated with a higher level of healthcare use. This result however, is not found for individuals and so it would appear that the characteristics of the household head are important for determining the use of healthcare services for the household rather than individual level characteristics (Gustafsson-Wright et al., 2011).

Using an OLS model, Gustafsson-Wright et al. (2011) examine the effects of different shocks on a variety of different economic outcomes; income, medical expenditure, labour supply, consumption, assets and credit. Insured and uninsured estimates are obtained separately following the approach of Wagstaff (2007). A dummy variable is used for the different health shocks and takes on a value of one when the shock occurs and zero for when it does not occur. In addition, other independent variables are added including household characteristics such as age, age squared and household size. Community characteristics are also included to account for local health conditions. The analysis undertaken is at the household level and a household is insured if at least one household member is registered in an insurance scheme. The initial analysis undertaken in *Chapter 6* follows a similar approach and is then developed with propensity score matching.

Gustafsson-Wright et al. (2011) conclude that the uninsured experience high medical costs despite having access to public health centres. Moreover, even with support from friends or family, the costs experienced by these households remain high and households require additional financial support to assist their recovery. Without this support, households seek to sell assets, reduce consumption or take up loans in order to alleviate the financial burden that they face following a health shock. Those households which are insured are much better protected against the risk of having to use these coping strategies. However, access to credit and insurance markets can be limited in some cases. Therefore, households may only be able to pursue informal avenues of support (Beegle, 2003; Gan et al., 2006). There is a question within the KIDS dataset regarding medical aid, which is a form of health insurance; however, this question is only answered if there has been at least one incidence of morbidity in the household in the 15 days prior to the survey. Hence, this indicator is not included within the subsequent analysis as current medical aid status may not be reflective of the shock period considered within the analysis which is 5-6 years. Moreover, unless there was an incidence of morbidity in the previous 15 days the response will be missing.

Donovan et al. (2003) examine the coping strategies used in the presence of morbidity and mortality in Rwanda. A key finding from their work is that households make adjustments in the most cost-effective way that they can; maintaining production on the farm is a top priority. Therefore this shapes the series of coping mechanisms that are employed. Yamano and Jayne (2004) also find a similar type of response; households seek to hold onto certain assets in favour of others. Donovan et al. (2003) use a series of datasets to examine the coping strategies used by 1,231 rural households in response to mortality and morbidity of prime aged adults, which has occurred within a household during the last four years. The time period used within the KIDS households is a little longer at 5 years (for the 1998 analysis) and 6 years (in the case of the 2004) which is driven by the question presented within the survey.

They define a prime-aged adult as someone aged 15-60 years of age. Moreover, they also define chronic illness as someone who has been unable to work for a minimum of three months within the last year. Fifteen percent of households in the sample experienced a death in the last four years and 5 percent experienced a prime age death

(Donovan et al., 2003). As previously noted, ideally I would like to characterise health shocks in some way to individuals but this is not possible given the presentation of the KIDS data. The characteristics of the deceased were concentrated amongst the household heads, spouse of the household head or their adult children. Just over a fifth of all adults who died were aged between 15-25 years of age and the average length of illness prior to death was 23 months (Donovan et al., 2003). Households were asked to provide details on how a death had affected agricultural and livestock activities, other sources of income, savings and debt, food intake and the effects on any children.

Donovan et al. (2003) find that in the presence of morbidity, the most commonly used coping strategy by other household members was to; increase their working time, share activities with other households or reduce leisure time. Other coping strategies used included reducing the amount of land farmed, leaving a portion of land to fallow (Donovan et al., 2003). This is not an area of exploration within *Chapter 6* but it provides evidence that households do undertake adjustment behaviour in order to minimise the impact of shocks to their welfare.

Slightly less than a fifth of all households indicated that there had been no negative effect upon the household following a period of ill health of at least one member. This finding suggests that the effects of morbidity across households vary with the types of activities the household is involved in and the length of illness experienced. At the other end of the spectrum two households indicated that they had lost all their land (Donovan et al., 2003). This evidence illustrates that the distribution of shocks can be very different across households and therefore it is important to consider whether there are specific household characteristics that are important for policy interventions; if health shocks have very little impact on households then such cases are not a primary concern; scare resources should be directed at the most vulnerable.

Janye et al. (2005) also examine the effects of mortality on rural households in Zambia. The difference in their approach is that they consider the community rather than the household as the unit of analysis as they believe that the effects of mortality will be transmitted across afflicted and non-afflicted households.[50] Jayne et al. (2005) argue that traditional approaches will not capture the indirect effects of mortality which may be transmitted across households.

Other households could be indirectly affected by mortality as they provide support to care for orphaned children, for example (Pitayanon et al., 1997). Moreover, Pityanon et al. (1997) also note that in cases such as HIV/AIDS, the effect is *"rippling through families, communities and regions".*[51] Therefore non-afflicted households are not a suitable control group for comparison as they may not be completely separated from afflicted households. Jayne et al (2005) use community level data of 5,420 households in 393 communities within Zambia between 2001 and 2004 to examine the effects of prime-aged mortality on; land cultivation, crop output and per capita income. A similar analysis is also undertaken by Yamano and Jayne (2004) in Kenya.

Jayne et al. (2005) consider 6,922 Zambian households which were initially surveyed in 2001 and 5,420 were re-interviewed in 2004. Just over a tenth of households experienced at least one prime-aged death in the three-year period between surveys and almost ten percent of these households experienced more than one death. Yamano and Janye (2004) use a two-year panel of 1,422 households between 1997 and 2000. Beegle (2003) also carries out another similar study in Tanzania using the Kagera Health and Development Survey (KHDS). Households were interviewed approximately four times between 1991 and 1994. The aim of the study was to examine the effect of mortality upon surviving household members in terms of their time allocated and range of farming activities.

Jayne et al. (2005) examine a Pearson correlation between mortality and HIV prevalence rates. They find that there is a relatively close relationship between the two given a Pearson correlation coefficient of 0.84. HIV prevalence rates were obtained using antenatal clinic data. Yamano and Jayne (2004) take a slightly different approach. They use a study by Urassa et al. (2001) which was able to examine mortality rates for Tanzanians in the Kisesa ward between 1994 and 1998. Blood tests were used to

[50] Where an afflicted household is defined as one which has experienced mortality within a particular period.
[51] p.3.

determine HIV status. Therefore, it is possible to estimate the number of deaths (and rate of mortality) that would be anticipated due to HIV related conditions and for non-HIV positive individuals. Yamano and Jayne (2004) then use this information broken down by gender and age to generate a baseline estimate of deaths for different groups and can estimate those which are attributable to HIV and those that are associated with other conditions. There is very limited data within the KIDS dataset on HIV prevalence and therefore it is not possible to explore the role of HIV as a health shock in any detail.

Using a difference-in-differences (DID) model, Jayne et al. (2005) find that the total area of land under cultivation is negatively related to mortality rates. As KIDS is a panel dataset it is possible to consider a DID approach for a subset of households. In the two most recent waves of KIDS (second and third wave; 1998 and 2004 respectively) new households were also included which represent those households which have developed as children have grown up and established their own household unit. Such households would not be included within a DID approach as they did not exist in 1998. DID is not however explored within this thesis; instead I opt to use propensity score matching as an alternative method to deal with possible selection bias and this also enables me to look at the effect of health shocks in both periods..

As in many other studies they find that education plays an important role in determining the fall in land cultivation; in areas where the level of education is higher, the fall in land area cultivated is greater. This is also true when considering income; the effect of prime age mortality is greatest when people are more educated. Wealthier communities appear to cope better which is a similar result to that found by Donovan et al. (2003). Households who experience better initial conditions such as being wealthier have greater access to savings, assets and other resources; and are able to employ a broader range of coping strategies than less wealthy households which is not a particularly surprising result.

Yamano and Jayne (2005) find that the death of a working-age adult is not associated with a fall in farm the land area under cultivation. Barnett and Blaikie (1992) find a similar result in their study of the Rakai district in Uganda. What Yamano and Jayne (2004) do find, is that, households which experience a death experience less growth in the area of land used for high value crops between 1997 and 2000. They suggest that men often take care of high value crops within a household. Therefore, the death of a working age male has significant effects upon the cropping decisions that households make. The difference between the area dedicated to high value crops experiencing no death and a working age female death is not statistically different from zero. Therefore, this suggests that the gender of the deceased can be very important for certain outcomes. As previously indicated, it is not possible to include an indicator which identifies the recipient of the shock more clearly which is a limitation of the analysis conducted in *Chapter 6.*

Yamano and Jayne (2004) also examine changes in household assets and off-farm income which are two possible methods by which households can seek to protect themselves against shocks as income and assets can be used to purchase additional health services to replenish the stock of health capital. They consider changes in the value of assets such as farm equipment and livestock over the survey periods. They find that households which experienced a death reduced their asset holdings. They also find that households which experienced a death report lower levels of off farm income and the reduction in off farm income is greatest following the death of a male household head. This finding is attributed to the fact that almost half of the male deaths that occurred during 1997 and 2000 were associated with males who were in the top income quartile. Moreover, splitting the sample into two groups; poor and non-poor based on asset holdings in 1997, they conclude that relatively poorer households are more severely affected by a male mortality relative to their less poor counterparts. However, there is no conclusive evidence in support of this finding if the death is of a female member. They argue that it is important to understand whether the results that are generated are sensitive to initial conditions as this has implications for policy intervention when resources are scarce, so that we can ensure we target the most vulnerable.

Pityanon et al. (1997) examine changes in asset holdings following the death of a household member as an alternative economic measure of health shocks. They suggest that the use of savings rather than the sale of assets will have the least damaging effect on future activities as households are able to maintain their productive resources such as farm equipment. However, they also argue that the use of savings could also hinder future investment opportunities for the household. Therefore Pityanon et al. (1997) find evidence that non-productive assets are sold first when savings cannot be drawn upon. The sale of assets is used by around a fifth of all households in response to the death of a household member from HIV/AIDS related causes; land is the most commonly sold asset. Moreover, whilst Pityanon et al. (1997) find evidence that the funds from the sale of assets are used to finance increased medical expenditure they also find that funds are used to finance current consumption. No coping strategy is ideal; it is a case of implementing the best case scenario.

Chapter 6 does not consider changes in asset holdings as an outcome primarily because income is the specific topic of interest. Another issue is that the information primarily utilised within the chapter is taken from the economic shocks module. Whilst there is a question about the value of items that are lost as a result of a shock this question is only asked in the case of the loss of property, crops, animals or vehicles. An alternative if a DID was being used would be to consider the difference in durable goods owned by the household across the two periods. Although, given the significant amount of time between the waves and the fact that such items are household durables, there could be substantial differences in asset holdings across the two surveys which may have very little to do with health shocks.

Gan et al. (2006) consider the impact of health shocks upon household income in China using data from 1986 to 2002. Their measure of health shocks is what they refer to as "abnormal increases in household medical expenditure". Again, such an indicator is not selected as being indicative of a health shock in *Chapter 6* as there is insufficient data available within KIDS. An interesting finding from this study is that, in contrast to Jayne et al. (2005), recovery times following a shock are considerable; 19 years. Such a long period of recovery is perhaps questionable; nonetheless recovery times may be quite substantial. Moreover, they find that shocked households follow an income path which is approximately 12 percent lower than what it would have been in the absence of a shock which suggests that the effect of health shocks on income is considerable.

2.5.4 Conclusion

This section highlights that there are a number of different choices of health shock that we may consider as well as differing economic outcomes to explore. The analysis explored in *Chapter 6* utilises the Grossman (1972) model of health as a framework; with shocks representing a reduction in the stock of health capital at the household level. A reduction in health capital is anticipated to lead to a reduction in healthy days which are available to undertake productive activity and thus lower income. Income is used as an input; to purchase health services. Therefore, if income is sufficiently reduced this may limit the investment that households can make in health capital over time and limit future opportunities.

Some of the studies reviewed in this section specifically consider the effects of HIV/AIDS related deaths, whilst others consider deaths more generally or other measures of health shocks such as hospitalisations or healthcare expenditures. The choice of health shock is often driven by the data available. Ideally a range of different indicators may be considered so that different aspects of health status can be explored. *Chapter 6* therefore seeks to consider health shocks which are directly reported in the new "economic shocks" module of the KIDS survey which includes an indicator for a serious injury/illness shock (which also includes an exploration of duration), death of a household member and changes to health status on a ranking scale.

Different studies also consider the impact of health shocks on different outcomes. This includes income, consumption, the use of healthcare facilities, and the uptake of insurance to reiterate a few. *Chapter 6* specifically examines income and also explores the use of self-reported income losses that are associated with health shocks as a complement to regression analysis.

What is very clear from all the studies is that certain groups appear to be far more vulnerable than others and therefore are less able to adapt to these shocks. If adaptation to shocks were costless then the effects of health shocks upon households would not be a concern. What has been demonstrated in a series of studies is that *some* households are very badly affected by health shocks and experience major losses of welfare. The analysis that follows therefore seeks to examine whether households in KwaZulu-Natal province fall into the latter category.

2.6 Conclusion

This literature review considers three distinctive themes; child health, working age health and household health shocks. The common thread through them all is the exploration of the relationship between health and income. Health affects economic outcomes and these economic outcomes such as income may affect health. The problem is disentangling these two issues from one another in practice and this has thus far proven to be somewhat elusive.

The development of the Mosley and Chen (1984) framework which is utilised in the empirical work undertaken in *Chapter 4* has enabled researchers and policy makers to think about the various determinants of health and the interactions these determinants have on different levels; institutional, economic, socio-cultural and environmental. This work has been built upon by Huynen et al. (2005) to consider distal and contextual determinants. Our understanding of the determinants of health is just one step forward in improving health outcomes for people.

In *Chapter 5* I consider the human capital literature using a standard Mincer (1974) type earnings equation that includes health as a form of human capital. The discussion surrounding how we might measure health status indicates that there is a range of health indicators that one may consider; each of which has its advantages and disadvantages. In practice, the measure chosen may be fully or at least partially determined by the data available. The measure of health we select can capture particular characteristics of health but this can be at the expense of ignoring other equally important considerations of health. Therefore, it is important that we have an awareness of what each indicator of health is actually measuring.

The final empirical chapter considers the Grossman (1972) model of health and explores whether households' income may be affected by "health shocks".

I will now move on to *Chapter 3* which presents a discussion of the two datasets used within this thesis.

Chapter 3: Background information and data for analysis

The empirical analysis undertaken in this thesis uses information obtained from two surveys, for Malawi and KwaZulu-Natal province in South Africa, respectively. My investigation into the determinants of child and infant mortality in *Chapter 4* and the effects of adult morbidity on wages in *Chapter 5* both use the Malawian Integrated Household Survey (IHS2) which is a nationally representative household survey. In *Chapter 6*, my analysis of the relationship between income and health outcomes uses the KwaZulu-Natal Income Dynamics Study (KIDS) which is a panel study of a random selection of households in the KwaZulu-Natal province of South Africa. *Chapter 6* uses the two most recent waves of this study, which were undertaken in 1998 and 2004.

Following the introduction to this chapter I will first discuss the IHS2 data which is related *to Chapters 4 and 5* which examine Malawi. This is followed by a discussion of the KIDS data which pertains to *Chapter 6.*

The First Integrated Household Survey (IHS1) was undertaken in Malawi by the National Statistics Office in Malawi with support from the World Bank, as part of the Living Standards Measurement Study (LSMS) during winter 1997 to autumn 1998. The Living Standards Measurement Study was established in 1980 by the Development Research Group within the World Bank. Two key features of the LSMS surveys are multi-topic questionnaires designed to examine different aspects of household welfare and behaviour, and quality control (Grosh and Glewwe, 1995).

Initially the LSMS project sought to evaluate previous household surveys so that they could learn from past experiences about what information was useful to have and how it could be used. From 1985 to 1991 the focus moved to the implementation of the LSMS surveys in different countries to provide analysis about the links between household behaviour, living standards, constraints faced by households and how government policy can shape such constraints.

The purpose of the IHS in Malawi was to assess different aspects of welfare in the country to ensure that there was a thorough understanding of the most vulnerable within Malawi. The survey was designed to cover a broad spectrum of subjects, primarily to provide a comprehensive overview of the issues affecting those households living in poverty. In order to achieve this goal, the survey specifically collects information from a nationally representative sample of the population across all regions, rural and urban areas. The key areas of interest within the IHS data are to assess household behaviour, welfare, income, employment, health and education (World Bank, 2006). The survey was followed up by a second survey during the spring of 2004-2005 and most recently the third IHS was undertaken in 2010 with the assistance of a number of large donors including the World Bank, DfID, the Norwegian Embassy, Irish Aid and the German Agency for Technical Cooperation (GTZ).[52]

Throughout *Chapters 4* and *5*, I use the second IHS (IHS2) for the purposes of my analysis. I decided to use the second IHS as it was the most up to date IHS available at the time of starting this thesis. The final empirical chapter presented in this thesis explores how health shocks may affect household income in the KwaZulu-Natal province in South Africa using a survey which was part of the national Project for Statistics on Living Standards and Development (PSLSD). Specifically, I use the KwaZulu-Natal Income Dynamics Study (KIDS) which is currently a three-wave panel survey beginning in 1993 and followed up in 1998 and 2004.

One of the main aims of the KIDS survey is to gain a better understanding of the impact of poverty reduction policies and the effect of the rise in HIV/AIDS infection rates. The most recent wave of the KIDS survey was undertaken in 2004 with support from the South African Department of Social Development (DSD), a number of departments with the University of KwaZulu-Natal and the University of Wisconsin-Madison, as well as the International Food Policy Research Institute (IFPRI), the London School of Hygiene and Tropical Medicine (LSHTM) and the Norwegian Institute for Urban and Regional Studies (NIBR).

The KIDS survey has a double pronged approach; there are both quantitative and qualitative modules within the survey designed to collect information on a variety of topics. These topics include information on household composition, expenditures on

[52] http://www.dfid.gov.uk/stories/case-studies/2010/measuring-poverty-in-malawi/

food, durable and non-durable goods, education, health, agricultural production, employment and other sources of household income. In the most recent wave, some additional topic areas were added; a specific topic of interest for the purpose of my thesis is the addition of economic shocks which are discussed in section 3.6.5.

The remainder of this chapter is as follows: *Section 3.1* will present some background information about Malawi. *Section 3.2* will outline how the IHS2 was undertaken. *Section 3.3* will outline some of the characteristics of the IHS2 data, transformations that took place and key modules of interest which are used within the analysis that takes place in *Chapters 4* and *5.*[53] I will then move on to a similar discussion of the KwaZulu-Natal province in South Africa in *Section 3.4*. This is followed by a more in-depth discussion of the KIDS dataset in *Section 3.5*. In *Section 3.6* I will present some of the key topics that will be further explored in *Chapter 6*. *Section 3.7* will conclude the chapter.

3.1 Background to Malawi

This section presents a brief background picture of Malawi, including a short overview of the health challenges that the country experiences.

Geography of Malawi

Formerly known as Nyasaland, the country gained its independence from British colonial rule in 1964 and renamed itself to the country we know today as Malawi. It is a landlocked country located in the southern part of Africa, sharing its largest border with Mozambique, followed by Zambia and Tanzania.

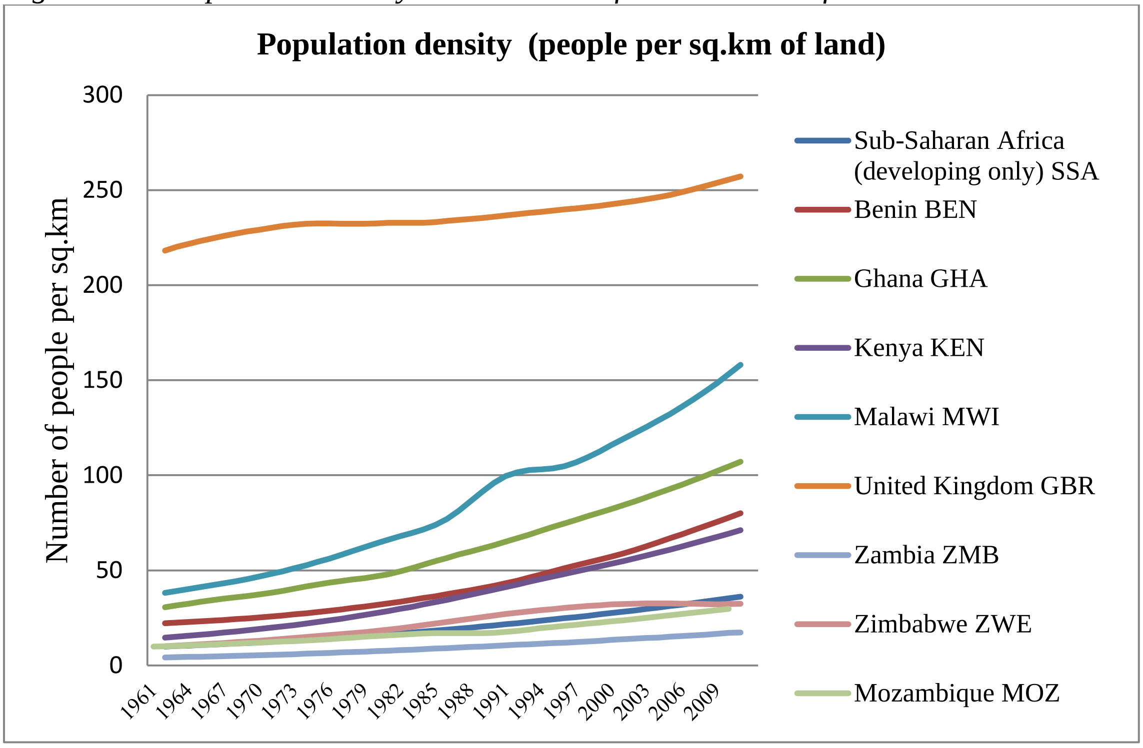
Around a fifth of the country is covered in water, primarily from Lake Malawi which is the third largest lake in the whole of Africa and runs through the middle of the Great Rift Valley. Malawi has a sub-tropical climate and experiences two main seasons, a rainy season which runs from November to May, followed by the dry season for the remainder of the year. The agricultural sector employs over 85% of the workforce,

[53] The descriptive statistics therefore specifically focus on the observations that were utilised across the two chapters.

however, it only contributes to just over a third of national income (Harrigan, 2005). Agricultural exports account for 90 percent of all foreign exchange earnings within the country.[54]

The Malawian population is continuing to grow, with a recent indication reporting that the total population is now 15.4 million[55] (World Bank, 2012). Annual population growth is just in excess of 3 percent per annum, which is significantly higher than the rate of growth within low income countries collectively at 2.1% per annum and 2.5% per annum in sub-Saharan Africa as a whole[56] (World Bank, 2012).

Figure 3.1.1 Population density in a selection of sub-Saharan African countries



Source: World Bank (2012)

Population density is also very high in Malawi compared to other countries within sub-Saharan Africa, as Figure 3.1.1 shows. The most recent estimate from the World Bank (2012)[57] indicates that average population density in Malawi is 158 people per square

[54] http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:22356350~menuPK:3266877~pagePK:51236175~piPK:437394~theSitePK:73154,00.html
[55] 2009 figure.
[56] 2009 figure.
[57] 2010 figure.

kilometre, compared to 80 people in Benin and 17 people in Zambia; for sub-Saharan Africa as a whole, there are 36 people per square kilometre.

The HIV prevalence rate within the total population in Malawi has been steadily declining since its peak at nearly 15% in the late 1990s, falling to 11% in 2009 (World Bank, 2012). There has also been a sustained improvement in the life expectancy at birth, increasing from 46 years for males and females in 1999 to 54 years for females and 53 years for males respectively by 2010 (World Bank, 2012). Whilst there have been a number of improvements in the key development indicators within Malawi, it remains, however, as one of the poorest countries in the world.

Health issues

Whilst HIV prevalence is high in Malawi it is not the only serious health issue the country faces. Malaria also affects a large proportion of the population. The World Health Organisation (WHO) indicates that funding for the prevention of malaria has been increasing, but there is still a funding shortage (WHO, 2012). A focus in the fight against malaria has been the encouragement of an increase in the use of insecticide treated nets (ITNs) and the number of households owning one in sub-Saharan Africa has risen dramatically from 3% in 2000 to 53% in 2011 (WHO, 2012). Moreover, it is also reported that the use of indoor sprays has also been increasing from 5% in 2005 to 11% in 2010 (WHO, 2012). There have also been improvements in access to testing and appropriate medication, particularly within sub-Saharan Africa (WHO, 2012).

The result of these steps has been a significant reduction in the reported number of malaria cases in 43 of the 99 countries which are classified as having on-going transmission. In these 43 countries the reported number of malaria cases has declined by over 50 percent and in another eight countries, there has been a reduction in the number of cases of between 25-50% (WHO, 2012). To put into context the scale of the problem within Africa, the WHO (2012) reported that 81% of all cases of malaria in 2010 were in the region and 91% of malaria associated deaths also occurred in the region. Moreover, malaria claims the lives of the young, with 86% of all malaria deaths worldwide accounted for by children aged under-five (WHO, 2012).

Malawi is classified by the WHO (2012) as a high transmission country, where there is at least one case of malaria per 1000 people. Steps taken to reduce the number of malaria cases include free insecticide net distribution which began in 2006, and this has been extended more recently to include all age groups as well as indoor spraying which was adopted in 2007 (WHO, 2012). Expenditure on testing and treatment accounts for a relatively small proportion of malaria expenditure in Malawi according to evidence presented in the WHO (2012) report. Moreover, the number of admissions and deaths associated with malaria do not appear to have improved since 2001 in Malawi (WHO, 2012), which indicates that whilst steps are being made to tackle malaria, it still poses a substantial risk to human health in the country.

3.2 The Second Integrated Household Survey (IHS2)

This section provides some information on how the data within the IHS2 was collected and verified for accuracy. In the next section a number of transformations to the data are outlined for the purposes of the empirical analysis which I undertake in *Chapters 4* and *5*. In total 51,297 household members across 11,280 households were included in the IHS2 across all three regions of the country. There are a total of 31 modules within the IHS2. The health module is answered by all individuals; with mothers or guardians reporting on behalf of children less than 10 years of age. Questions about education are presented to all individuals aged over five. Time use questions are directed to all individuals aged over five. Questions related to security and safety are also presented to all individuals greater than 10 years of age. Questions surrounding housing, food, expenditure, durable goods, enterprises, other sources of income, gifts, social security nets, credit, assessment of subjective well-being, shocks and deaths are presented to the household head. Child anthropometry questions are only considered for children aged between 6-60 months of age and are answered by the mother or guardian. Questions that are associated with the different types of agriculture are directed at all individuals who identify that they are engaged in that particular activity.

A household member within the IHS2 is automatically defined as someone who has been living in the household for a minimum of three months in the previous 12 months. Therefore a visitor who has been staying in the household for more than three months is counted as a household member. Servants, hired workers and lodgers are not considered to be household members if they have their own household elsewhere that they are dependent upon, or are the household head for. Moreover, the following types of individual are still classified as a household member even if they have been living in the household for less than three months: household head, infants less than three months, new spouses who are now living in the household, and those who may be living somewhere else but are dependent on the household; for example a child living in a boarding school.

Of the households initially selected, 96 percent (10,777 households) were available for interview upon being visited by the enumerator (NSO, 2005). An additional 507 households were selected as replacements as the original household could not be interviewed (NSO, 2005). The reasons for replacement are summarised in Table 3.2.1 below.

Table 3.2.1 Reasons for replacing household from original sample selection

Reason for replacement	Number of households replaced for given reason
Dwelling found but no household member could be found	197
Dwelling found but respondent refused	41
Dwelling found but appears unoccupied	180
Dwelling found but not a residential building	12
Dwelling destroyed	43
Dwelling not found	30
Total households replaced	507

Source: NSO (2005)

Households were selected using a two-stage stratified approach; the first stage was to select the enumeration areas (EA) from the 1998 Population and Housing Census and in the second stage twenty households were randomly selected from each of the EAs which cover the thirty districts of Malawi; twenty-six rural and four urban (NSO, 2005). Enumerators were then assigned to survey the 20 randomly selected households from a list of all eligible[58] households in each of the EA per month. A household is defined as

[58] People living in a private dwelling. Therefore prisons, hospital and army barracks were excluded. Members of the armed forces living in barracks were excluded (but not those living in private residences). Non-Malawian diplomats, staff and members of their household are also excluded. Non-Malawian citizens who are not diplomats or diplomatic staff are included if they live in a private residence. Tourists were also not included.

an individual or group of people (that can be related or unrelated) that live together as a single unit. This is defined as having “common housekeeping arrangements” in the enumerator manual. Such arrangements include living together, pooling money and eating at least one meal a day together. Families and households in this context may not be the same; a household is an economic unit whereas a family describes relationships. In practice they may often be the same but it is not necessarily true; and therefore families living in the same home but without common housekeeping arrangements would represent separate households according to the enumerator manual.

The enumerator was responsible for finding the listed households and conducting the interview for the survey. If the household could not be selected then a supervisor could authorise that one of the five replacement households could be selected. There is no evidence within the enumerator, field supervisor or zone supervisor manuals that incentives were provided to households to aid completion of the survey.

The sample selection for each district is summarised in Table 3.2.2 below. In addition to the modules included within IHS1, in IHS2 modules were added to examine security and safety, social safety nets, credit, subjective assessment of well-being and recent shocks to the household. Two particular modules of interest to me are the final two additions listed above which I will discuss in more detail later on in this chapter in sections 3.3.8 and 3.3.9. Prior to the survey being conducted eighty trainees were recruited and underwent four weeks of training. Upon completion of the training period 15 trainees were appointed as field supervisors, 47 became enumerators, 12 became data entry clerks and 10 trainees were kept in reserve (NSO, 2005). In addition a further twelve members of the NSO were trained to act as zone supervisors for the eight zones[59] from which the fieldwork was organised.

The fieldwork started in March of 2004 across all areas of the country and was completed by April 2005. The aim of the survey was to ensure that enumerators did not need to revisit households and the whole survey could be conducted in one visit, so seasonal factors were also taken into consideration (NSO, 2005). In order to carry out

[59] The eight zones were; Karonga, Mzuzu, Kasunga, Salima, Lilongwe, Blantyre and Ngabu within the Agricultural Development Division (ADD) headquarters (NSO, 2005).

the fieldwork, each zone was allocated one zone supervisor, some field supervisors and enumerators, a vehicle and a driver (NSO, 2005).

3.2.1 Quality control

Table 3.2.2 Sample selection for the IHS2

District	HH population projection	EAs	IHS2 sample HHs
Chitipa	31,006	12	240
Karonga	47,147	12	240
Nkhata Bay	38,062	12	240
Rumphi	28,849	12	240
Mzimba	108,046	24	480
Mzuzu City	27,144	12	240
Kasungu	118,607	24	480
Nkhotakota	60,007	12	240
Ntchisi	44,537	12	240
Dowa	103,239	24	480
Salima	72,787	12	240
Lilongwe	251,640	48	960
Lilongwe City	141,389	24	480
Mchinji	86,092	12	240
Dedza	135,849	24	480
Ntcheu	101,707	24	480
Mangochi	176,345	36	720
Machinga	101,839	24	480
Zomba	139,810	24	480
Zomba Urban	21,719	12	240
Chiradzulu	67,912	12	240
Blantyre	85,110	12	240
Blantyre City	163,393	24	480
Mwanza	37,941	12	240
Thyolo	131,835	24	480
Mulanje	122,974	24	480
Phalombe	71,573	12	240
Chikwawa	94,237	24	480
Nsanje	49,817	12	240
Balaka	70,732	12	240
Total	*2,731,346*	*564*	*11,280*

Source: NSO (2005)

In addition to the supervisors at the field and zone level, additional checks to ensure the quality of the data were also carried out on a monthly basis during visits by members of the core management team for the IHS2 (NSO, 2005). Whilst data was being collected, the data entry team also began processing the data that had been collected using the Census Survey and Processing System (CSPRO) software.

The CSPRO software is a free piece of software which was developed by the United States Census Bureau, Macro International and Serpro, S.A and is designed for use in entering information from a census or other similar type of survey (USAID, 2010). The advantage of this particular software is that it is able to provide some preliminary checks on the data, such as examining the range of reported values and cross referencing between different modules across the survey questionnaire (NSO, 2005).Following the initial checks which the CSPRO software is able to process, additional manual checks took place to check for any errors (NSO, 2005).

3.3. The IHS2 indicators

In this section, I will examine some of the data modules within the IHS2 in additional detail, with particular reference to some of the key variables that are used within *Chapter 4* and *Chapter 5*. A number of variables were transformed for the purposes of empirical analysis and I provide an outline of the changes that took place in order to get the data into a tractable format.

3.3.1 Household location and services

Malawi has a very low rate of urbanisation as the majority of households within the sample are located within the rural area and fewer than 13 percent of all individuals are located in the urban area. An examination of only those cases that are subsequently utilised in *Chapter 4, and 5*, which totals 14,772 individuals[60] who are; workers, mothers or ganyu labour[61], indicates that almost half live in the Southern region. The overall proportions are very similar for the full sample of just over 52,000 individuals and the subsample. Twelve percent of my total IHS2 subsample utilised within *Chapter 4* and 5 live in the urban area, with the remaining 88 percent living in the urban area as shown in Table 3.3.2.

[60] Located across 8.597 households.
[61] An individual could have a combination of these characteristics.

Table 3.3.1 Location of individuals (full sample)

Variable	Mean	Standard deviation	N	Missing
Residence (rural/urban)	0.877	0.328	51293	4
Northern region	0.158	0.365	51297	0
Central region	0.397	0.489	51297	0
Southern region	0.445	0.497	51297	0

Source: IHS2.

The proportion of overall urban residences in each of the three regions varies a little, with 14.49 percent of individuals in the Northern region classified as being in the urban area, followed by 13.39 percent of households in the Southern region. The central region has the lowest percentage of households located in the urban area at 10.10 percent.

Table 3.3.2 Location of individuals (sub sample)

Variable	Mean	Standard deviation	N	Missing
Residence (rural/urban)	0.880	0.325	14772	0
Northern region	0.118	0.322	14772	0
Central region	0.402	0.490	14772	0
Southern region	0.480	0.500	14772	0

Source: IHS2

The vast majority of individuals live in homes that do not have a permanent roof structure or any kind of floor covering. The most common source of water is from a pump or protected spring, and most individuals have access to a latrine rather than a flushing toilet. The descriptive statistics for the subsample are very similar and have not been reported here as specific details on sanitation that are pertinent to *Chapter 4* are included within that chapter.

3.3.2 Household roster

In addition to some general characteristics about the household, there are also a number of modules which seek to gather information about individual household members. A key module, which provides some additional detail about some of the demographic characteristics of the household members within the survey, is the household roster.This module provides information about each individual who is classified as a member of the household as previously outlined. Table 3.3.4 provides an overview of the sample by age group, gender and location, whilst Figure 3.3.1 illustrates the age distribution by gender using a population pyramid.

Table 3.3.3 Living conditions all individuals (full sample)

Variable	Mean	Standard deviation	N	Missing
Roof material				
Grass	0.723	0.448	51297	0
Iron sheets	0.265	0.442	51297	0
Clay tiles	0.008	0.089	51297	0
Concrete	0.0003	0.018	51297	0
Plastic sheeting	0.003	0.056	51297	0
Other	0.0001	0.010	51297	0
Flooring				
Sand	0.029	0.168	51283	14
Mud	0.756	0.430	51283	14
Cement	0.212	0.408	51283	14
Wood	0.0003	0.016	51238	14
Tile	0.003	0.053	51238	14
Other	0.0003	0.017	51283	14
Water				
Piped water	0.211	0.407	51927	0
Pump/protected spring	0.462	0.500	51297	0
Unprotected well/spring/reservoir	0.325	0.468	51297	0
Sanitation				
Flushing toilet	0.031	0.174	51297	0
Latrine	0.821	0.383	51297	0
No toilet	0.146	0.353	51297	0
Rubbish disposal				
Bin	0.029	0.167	51297	0
Pit	0.591	0.491	51297	0
Heap	0.179	0.383	51297	0
No rubbish	0.171	0.377	51297	0

Source: IHS2.

Overall the sample is well balanced across the different genders, with similar proportions of males and females included within each age group and location.

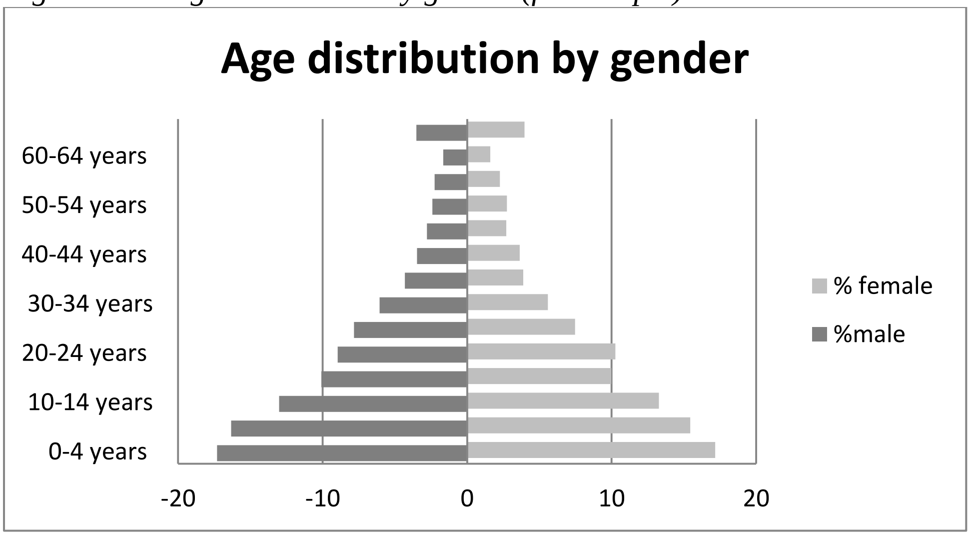
Across the entire sample, household size varies quite substantially and is broadly similar across the subset. Individuals used in the subset are generally older, as the subset excludes anyone under the age of 5 (the youngest age that someone could be engaged in working activities, whether formal or ganyu).

Table 3.3.4 Distribution of gender and residence by age (full sample)

Age group	Gender		Place of residence	
	Male	Female	Urban	Rural
0-4	17.3	17.2	15.6	17.5
5-9	16.3	15.4	13.0	16.3
10-14	13.0	13.3	12.5	13.2
15-19	10.1	10.0	11.1	9.9
20-24	9.0	10.3	12.6	9.2
25-29	7.8	7.5	11.7	7.1
30-34	6.1	5.6	6.8	5.7
35-39	4.3	3.9	4.6	4.0
40-44	3.5	3.6	3.7	3.5
45-49	2.8	2.7	2.9	2.7
50-54	2.4	2.8	1.5	2.7
55-59	2.3	2.3	1.5	2.4
60-64	1.7	1.6	1.1	1.7
65+	3.5	4.0	1.4	4.1
Total	100.00	100.0	100.0	100.0

Source: IHS2

Figure 3.3.1 Age distribution by gender (full sample)



Source: IHS2

Table 3.3.5 Household composition (full sample)

Variable	Mean	Median	Min	Max	Standard deviation	N	Missing
Age	21.410	16	0	115	18.739	51191	106
Gender	0.491				0.500	51290	7
Size	5.746	5	1	27	2.562	51293	4

Source: IHS2

Table 3.3.6 Household composition (subsample)

Variable	Mean	Median	Min	Max	Standard deviation	N	Missing
Age	30.069	27	5	97	13.688	14770	2
Gender	0.512				0.500	14772	0
Size	5.217	5	1	21	2.399	14772	0

Source: IHS2

Within the entire IHS2, the split between the genders in the sample is roughly even, with 49.1 percent of the sample being male and 50.9 percent who are female. Just over three-quarters of households have a male household head and 22.9 percent of households are headed by a female.

Date of birth is reported for each individual rather than age. The IHS2 team have then calculated age based on the difference between the date on which the survey took place and the date of birth for each individual as there is no evidence in the survey questionnaire that individuals were ever asked to report their age. Age is reported in months and years. There are a very small number of cases (106 in the full sample) where the age of the individual has not been calculated.

There is reason to suggest that the year of birth for each individual may not be accurately reported for a small number of individuals who are far older than one might anticipate given that the life expectancy in Malawi is less than 55 years of age for both men and women. A closer inspection of the reported ages of all individuals within the IHS2 reveals that the oldest person is reported as being 115 years old and just under one percent of all respondents indicated that they were 78 years old or more. Whilst it is feasible that there may be a small proportion of very elderly people in Malawi, I suspect that it is unlikely that there are people who are in excess of a hundred years old. As

Figure 3.3.1 shows, Malawi has a relatively young population based on the sample within the IHS2.

The mean and median age of all individuals within the IHS2 is 21 and 16 years of age respectively. Turning to just household heads for the full sample; it is observed that the mean age of a male head of household is 41 years with a slightly lower median age of 36 years reported. A female head of household tends to be slightly older on average, with a reported mean and median age of 48 years.

3.3.3 Education

Every household member aged five years or over is asked to provide details about their education. This includes information about what languages they speak at home, their ability to read and write in Chewa and English, and information about their schooling. Table 3.3.7 shows the different types of languages that are spoken within Malawian households used within the subsequent sub-sample

Many of the languages spoken within Malawi are part of a group of languages which is related to what is known as Bantu and this collection of similar languages is spoken in a large part of Africa from south of Nigeria across the Central African Republic and Democratic Republic of Congo, Uganda, Kenya and Somalia (Nurse and Phillipson, 2003). It is not uncommon for people within Africa to be bi- or multilingual according to Nurse and Phillipson (2003), with many Africans learning a local language at first, followed by a second national language and then finally an international language. Moreover, it is not unusual for people to be able to communicate in a number of different languages, their primary language for communicating with people in their own local area, and then other local languages which will assist them to communicate with people in the neighbouring areas (Nurse and Phillipson, 2003).

As Table 3.3.2 shows, a wide range of languages are spoken within Malawi, each of which has Bantu origins. Chichewa, which is also known as Chewa, is one of the two official languages spoken within Malawi, and is the most commonly spoken Bantu language within the Southern and Central regions of Malawi.[62]

For the purposes of the subsequent empirical analysis that I undertake in *Chapter 4* and *Chapter 5*, I have created dummy variables for the levels of education that individuals have obtained. Dummy variables for no primary, primary, secondary and tertiary education are generated where D=1 is indicative of the individual possessing that characteristic and D=0 where they do not. Individuals report their highest level of educational qualification. If no level of education was indicated at all (including no education), values are treated as missing.

Table 3.3.7 Language spoken by Malawian individuals (subsample)

Variable	Mean	Standard deviation	N	Missing
Chewa	0.578	0.494	14755	17
Nyanja	0.121	0.327	14755	17
Yao	0.096	0.295	14755	17
Tumbuka	0.087	0.238	14755	17
Lomwe	0.033	0.178	14755	17
Nkhonde	0.006	0.076	14755	17
Ngoni	0.008	0.087	14755	17
Sena	0.032	0.176	14755	17
Nyakyusa	0.001	0.030	14755	17
Tonga	0.010	0.100	14755	17
Lambya	0.009	0.067	14755	17
Senga	0.0001	0.008	14755	17
Sukwa	0.006	0.076	14755	17
English	0.002	0.041	14755	17
Other	0.010	0.100	14755	17

Source: IHS2

Where the individual has not managed to achieve even a Primary School Leaving Certificate (PSLC) as their highest level of education, they report that they have no primary. If the individual has managed to achieve a Malawi Junior Certificate Examination (JCE) or a Malawi School Certificate of Education (MSCE) they will be coded as D=1 and D=0 for each of the other dummy variables for education level. Finally, if the individual has been awarded a non-university diploma, university diploma or university degree at undergraduate or postgraduate level I classify them as having

[62] http://www.peacecorpsmalawi.org/malawi/

achieved a tertiary level of education and they are coded as D=1 for tertiary and D=0 for the other categories.

As can be seen in Table 3.3.8, the majority of respondents who answered the question of education within the subsample indicated that they had not achieved any level of formal education. This table highlights the general lack of formal education within the country. In *Chapter 5* the returns to education are examined; the above table highlights that even if the returns to particular levels of education are very high, very few people possess such levels of education. *Chapter 4* and *5* contain additional detail on the education of the sample by gender and location.

Table 3.3.8 Education of individuals (subsample)

Variable	Mean	Standard deviation	N	Missing
No primary	0.731	0.443	11238	3478[63]
Primary	0.109	0.312	11238	3478
Secondary	0.143	0.350	11238	3478
Tertiary	0.016	0.127	11238	3478
Rural				
No primary	0.788	0.409	9608	3388[64]
Primary	0.103	0.304	9608	3388
Secondary	0.103	0.304	9608	3388
Tertiary	0.005	0.073	9608	3388
Urban				
No primary	0.397	0.489	1630	146[65]
Primary	0.145	0.351	1630	146
Secondary	0.378	0.485	1630	146
Tertiary	0.081	0.272	1630	146

Source: IHS2

Around one fifth of the sub-sample indicated that they had never attended school. The primary reason for not attending school was reported to be a lack of money for fees or uniform, followed by being orphaned and experiencing a disability or illness which prevented them from attending school. Across all levels of qualification, a greater percentage of individuals have achieved a formal qualification in the urban area in comparison to the rural area. Moreover, a significantly smaller percentage of individuals

[63] The total population for the subsample is 14,772 individuals.
[64] The rural population for the subsample is 12,996 individuals.
[65] The urban population for the subsample is 1,776 individuals.

have no formal qualification in the urban area (39.7 percent) in comparison to just under 79 percent of individuals who reside in the rural area.

3.3.4 Health

Each household member was asked about various aspects of their health and the impact that any ailments had on their day-to-day activities. For children aged under 10 years of age, the mother of the child responded on their behalf.

An initial question in the health module seeks to establish whether individuals have recently experienced any bouts of illness. The response rate to this question is high, at ninety-nine percent for the full sample, and of those who responded seventy-two percent of individuals or their representative indicated that they had not experienced any illness or injury in the two weeks preceding the survey.

Table 3.3.9 Incidence of recent morbidity (subsample)

Experienced morbidity	Frequency	Percent
Yes	4,217	28.63
No	10,511	71.37
Total	14,728[66]	100.00

Source: IHS2

Taking a closer look at those individuals who indicated that they had experienced some sort of illness or injury in the last two weeks within the subsample, only 4 individuals did not provide an indication of the type of ailment they had experienced. The most commonly reported illness or injury was malaria, by 26.66% of respondents. Lower respiratory conditions were reported in 16.64% of cases, 5.86% experienced some sort of stomach complaint and headaches were also indicated in fifteen percent of cases. A wide range of other ailments were also reported, but not in any great frequency compared to those just listed above. Where individuals have experienced morbidity, there do not appear to be any predominant differences in the prevalence rate of lower respiratory conditions between rural and urban residences. However, there is a higher incidence of malaria in the urban area, with 36 percent of urban individuals who have

[66] Missing in 44 cases of the subsample.

recently experienced illness reporting that malaria is the cause of their health ailment, compared to 26 percent their rural counterparts.

Self-diagnosis of health conditions is common, followed by household members; medical professionals confirm health conditions in only a small proportion of cases (both for the full sample and the subsample). Even in serious illness cases where one would require formal medical treatment, the vast majority of illnesses remain diagnosed by the individual themselves or by another household member. However, despite not receiving a formal diagnosis in many cases, this did not prevent individuals from seeking some form of relief for their condition. In a large number of cases, medicine from a local shop can provide the relief required. Treatment at a health facility is the second most commonly sought out assistance and in many cases if the illness or injury is not serious, no action was required.

Table 3.3.10 Disability status in IHS2(subsample)

Variable	Mean	Standard deviation	N	Missing[67]
Disabled	0.177	0.381	12766	1962

Source: IHS2

Within the IHS2 no specific reference is made to the definition of "normal activities" and therefore I will make the assumption that this means people are still able to attend school or go to work and any other activities that one would normally do on a day-to-day basis. In *Chapter 5*, I use the information about whether or not someone could carry out their normal activities, together with the number of days that they were unable to carry out such activities, in order to define what I refer to as a disabled day which I use as a measure of morbidity. Whether or not the individual is able to continue their normal activities is subjective and determined purely by the respondent; what one person may consider to be debilitating and therefore stop their normal activities may not be the same for another individual who perhaps has the same ailment. In *Chapter 2*, I discussed some of the difficulties associated with using self-reported measures of health status. For the purpose of this chapter, I just present some of the basic descriptive statistics related to individuals' perception of their morbidity.

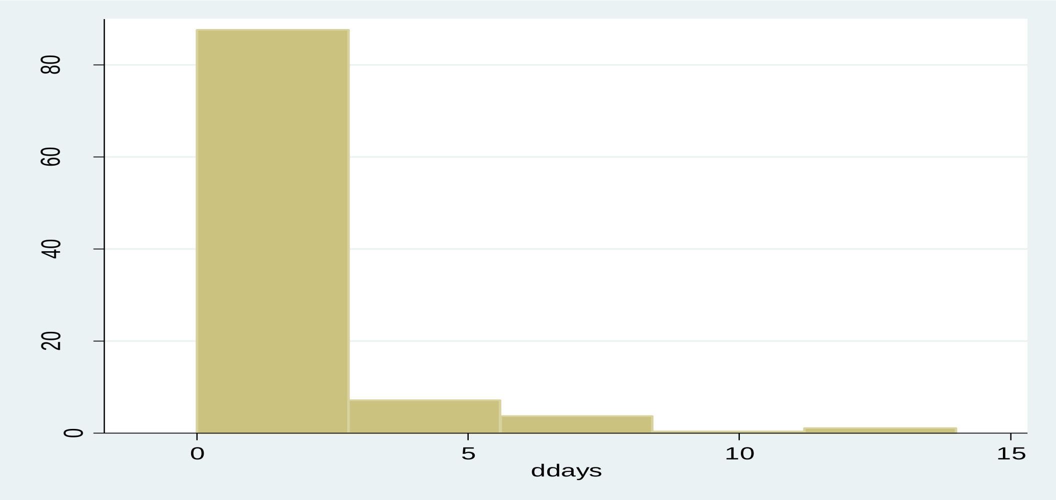
[67] Disability status is only reported if someone had experienced morbidity in the last 2 weeks. Recent morbidity was reported by 14,728 of the 14,772 individuals in the subsample.

In order to examine morbidity, I define an individual who has experienced an illness or injury in the past two weeks and who has had to stop their normal activities as someone who is "disabled". Clearly any individual who did not experience any illness or injury in the two week period prior to the survey cannot experience disability by default. Furthermore, someone who has experienced illness or injury may not have had to stop their normal activities, and therefore they would also not be described as disabled using the definition I have outlined. Therefore I can use a dummy variable where D=1 when the individual has experienced disability and D=0 when they have not experienced disability. Table 3.3.10 shows that the majority of those in the subsample did not experience any form of disability, either because they were in good health or their illness was not severe enough to become debilitating. Of those who did experience disability from this subsample, we can also examine the number of disabled days that they experience, defined as the number of days that they are unable to carry out their normal activities. The maximum possible number of disabled days any individual can experience given this definition is 14 days, as we are constrained to a two-week reference period. This is not to say that they did not experience disability for a longer period of time but I am unable to examine this further due to the questions asked within the IHS2. As Figure 3.3.2 shows, most individuals who experienced disabled days only stopped their activities for a few days.

In addition to questions about recent health experiences, individuals were also asked about their perception of health status compared to the previous year, the ease with which they could carry out activities such as sweeping the floor and walking five kilometres and whether they suffered from any form of chronic health condition. I do not explore these within the discussion, however, as they were not subsequently utilised within the empirical chapters.

Just fewer than twelve percent of the subsample individuals report that they suffer from some type of chronic illness. Twenty-six different conditions are reported in total. The most frequently reported chronic illness is arthritis/rheumatism by a little over a fifth of individuals within the subsample who experience a chronic illness, closely followed by asthma for almost 12.46 percent of individuals in the subsample, and malaria/fever is reported in 14.53 percent of cases. Perhaps not surprisingly, HIV/AIDS is reported by fewer than one percent of respondents. Given that HIV prevalence in Malawi is reasonably high, one might expect that more individuals are affected by HIV. However, very often people may be unaware of their HIV status, they may not wish to report their HIV status, or it may be that another health condition is more symptomatic and therefore that is the condition, which they feel is responsible for making them chronically ill.

Figure 3.3.2 Disabled days experienced by individuals in the two weeks prior to the IHS2(subsample)

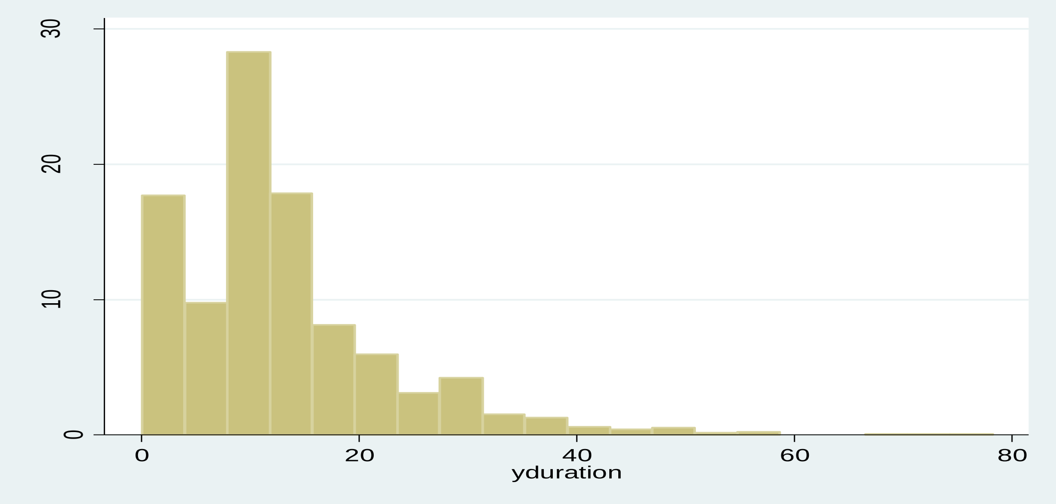


Source: IHS2

Figure 3.3.3 shows that just chronic illnesses may be experienced by individuals over a long period of time; such conditions may be more serious than others, some may enable adjustment and some may be debilitating, which is the area of exploration in *Chapter 5*.

With respect to perception of health, 35 percent of all individuals from the full sample indicating that their health now is much better than what it was a year ago and 19 percent indicated it is somewhat better in comparison to a year ago. A much smaller percentage of individuals indicate that their health has declined in the last year. Specifically considering the subsample, Table 3.3.11 highlights that just over a third of individuals within the subsample perceive that there health is better than a year ago and a relatively small proportion feel that it has declined.

Figure 3.3.3 Length of time (years) that the individual has been chronically ill (subsample)



Source: IHS2

There are a number of other questions which examine current health, such as whether the individual has a cough and the ability to carry out certain activities such as sweeping a floor. I do not present any further discussion of these indicators as I have not used them within my empirical work.

Table 3.3.11 Perception of health today in comparison to a year ago (subsample)

Variable	Mean	Standard deviation	N	Missing
Much better	0.363	0.480	14691	81
Somewhat better	0.186	0.389	14961	81
About the same	0.357	0.479	14691	81
Somewhat worse	0.088	0.283	14691	81
Much worse	0.006	0.075	14691	81

Source: IHS2

An important set of indicators within the IHS2 health module is related to recent pregnancies which I use to select my sample for *Chapter 4*. All women aged 12-49 years of age are asked whether they have given birth in the past two years, including any births that were stillborn. Further discussion of the specific sample selection and the details associated with the births can be found in *Chapter 4.*

3.3.5 Time use and labour

All household members aged five and over were asked about the types of general household activities they participate in, such as how long they spent collecting water and the types of paid employment activities they were recently involved in. For the purposes of this thesis I am going to concentrate on paid employment activities rather than more general household activities, as I specifically use information about employment within *Chapter 5*.

Household members were asked whether or not they have participated in ganyu labour in the last week and if so how many hours they worked. Ganyu labour can be thought of as *"any off-own-farm work done by rural people on a casual basis"* (Whiteside, 2000, p. 1). Whiteside (2000) indicates that ganyu labour may be very short term for only a few days, or it may be somewhat longer and cover several weeks. Payment can be in cash but equally it is common for payment to be in-kind (as food for example). Ganyu labour is also not constrained to only being on other small farms; it may also be used by large estate farms or in some cases in another neighbouring country (Whiteside, 2000).

As will be explored in *Chapter 5*, the amount of time people work in this sector is variable. Of the individuals who had participated in some form of waged employment during the last year and indicated the type of occupation they were employed in, the more frequently reported occupations included labouring and construction, security, agricultural based jobs, housekeeping and teaching. Additional aspects of working life are considered within *Chapter 5* and to some degree in *Chapter 4* as an indicator for labour force participation is included within the presented analysis.

3.3.6 Household ownership of durable goods

Information about the ownership of a number of different household assets is reported by each of the household heads. There are several categories of household assets which households report information about including: furniture, kitchen appliances, transport and agricultural inputs. In addition to indicating whether or not they own the particular asset they are also asked how many of each asset they own, how old it is and if they were to sell it, and what they think they would receive for it.

Table 3.3.12 presents information about the ownership of a selection of household assets for all households. In addition to those listed, households were also asked about a number of other assets such as televisions and washing machines. Other than bicycles very few households have any form of transportation such as access to a motorcycle/scooter, car, mini-bus or lorry. Road density in Malawi is relatively low, based on the most recently available estimate;[68] there are only thirteen kilometres of road per 100 square kilometres of land with only forty-five percent of these being paved across the 15,451km road network (World Bank, 2012). The road network within Malawi is important, as roads are the primary source of transportation for goods and people, but just over a quarter are in poor condition (The Roads Authority 2011). Therefore, the current five-year plan (2011-2016) from the Roads Authority seeks to make a commitment to reducing transport costs and improving access.

Table 3.3.12 Household asset ownership (full sample)

Variable	Mean	Standard deviation	N	Missing
Own sickle	0.550	0.497	11280	0
Own axe	0.633	0.482	11280	0
Own hoe	0.909	0.287	11280	0
Own panga	0.621	0.485	11280	0
Own clock	0.199	0.399	11280	0
Own lantern	0.304	0.460	11280	0
Own bicycle	0.362	0.480	11280	0
Own radio	0.549	0.497	11280	0

Source: IHS2

More widespread road access is particularly important for a country such as Malawi, which needs to get its goods to ports for export. In 1999, the Road Development Corporation was awarded the contract for Malawi's rail network. The aim has been to improve the routes through which goods can be transported both within Malawi and to and from its neighbours. A significant step forward has been a deal which was signed in early 2012 to build a new railway line through the country and to regenerate the older lines which had fallen into disrepair.[69] The new line will run between Blantyre in

[68] 2003 figure.
[69] http://www.railwaysafrica.com/blog/2012/02/new-line-to-malawi/

Malawi and Moatize in Mozambique and on to the port in Nacala. As Malawi is primarily an agricultural-based country, it is not surprising that the ownership of assets which would be used in agricultural production is generally relatively high, as shown in Table 3.3.12 and Table 3.3.13.

Table 3.3.13 Household asset ownership (subset)

Variable	Mean	Standard deviation	N	Missing
Own sickle	0.536	0.499	8597	0
Own axe	0.609	0.488	8597	0
Own hoe	0.907	0.291	8597	0
Own panga	0.618	0.486	8597	0
Own clock	0.203	0.401	8597	0
Own lantern	0.303	0.460	8597	0
Own bicycle	0.362	0.481	8597	0
Own radio	0.559	0.496	8597	0

Source: IHS2

3.3.7 Agricultural activity

Eighty-nine percent of households in the subsample[70] indicate that they engage in some form of agricultural activity or own a piece of agricultural land, although during the last cropping season only 512 households indicated that they rented out any land for others to cultivate.[71] During the last cropping season, a little more than a quarter of all households left some of their land uncultivated. However, in comparison to previous cropping seasons, the majority of households either maintained the amount of land that they used for farming in nearly two-thirds of all cases or increased the amount of land used for cultivation in just under a fifth of cases.[72]

During the last cropping season almost ten thousand households participated in rain-fed agriculture from the full sample, but just over half of these households received no revenue from their activities as they sold very little of their crop. This is not unexpected, as the majority of households have very small landholdings. Therefore, the likelihood is that any food crops grown will be for self-consumption rather than for sale. Far fewer

[70] The proportion is the same for the full sample of 11,280 households.
[71] These 512 households appear in both the full sample and the subsample.
[72] These proportions do not differ by the sample size particularly.

households participated in dry season cultivation during the last cropping season and again approximately two-fifths of all households did not gain any revenue from their crops which is indicative that they have been used for self-consumption. In addition to the food crops which are grown in the two main agricultural seasons in Malawi, income can also be generated from tobacco production, animal products and other forms of agriculture such as tree crops. Only 485 households from the full sample engaged in any form of animal husbandry and only 445 households from the full sample grew any tree crops, therefore I do not discuss these forms of agricultural production in any additional detail. Tobacco production is also relatively uncommon but I would like to provide a little more information about this for the reader.

Table 3.3.14 Tobacco earnings: net sales minus costs (subsample)

Variable	Mean	Median	Min	Max	Standard deviation	N	Missing [73]
Tobacco earnings MK	19.134.93	8600	-38000	800000	40966.36	1078	473

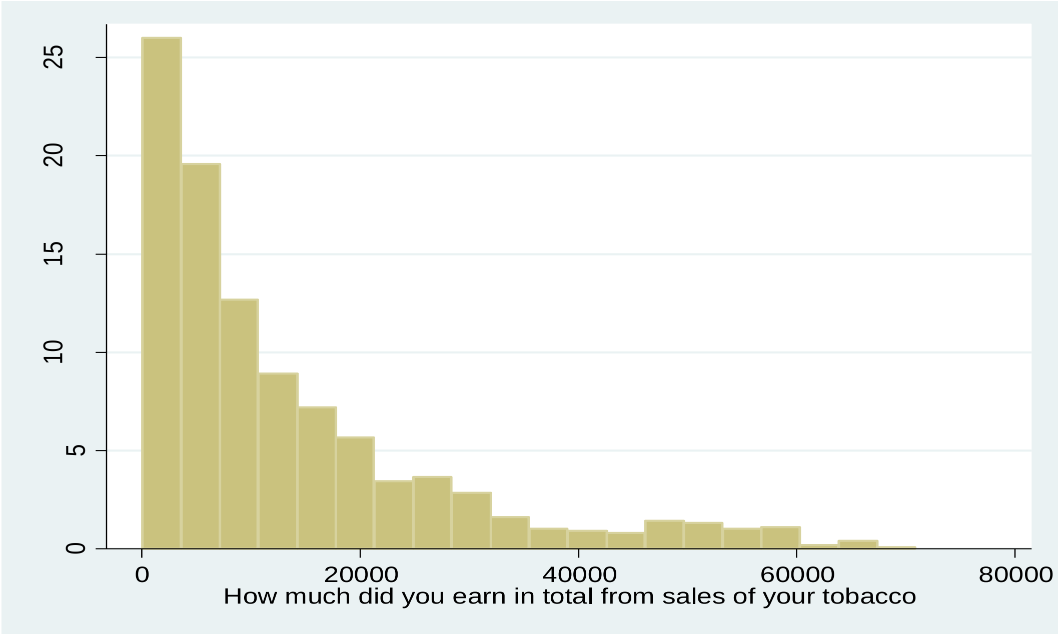
Source: IHS2

The majority of households engaged in tobacco production report their net income from the last season. For the subsample, there are some households which report very large earnings, ninety-five percent of households report earnings of 71,000 MK[74] or less. Figure 3.3.4 shows the distribution of net income from tobacco sales reported by the subsample of households with the exception of the top five-percent who reported earnings in excess of 71,000 MK and up to 800,000 MK.

During the last five cropping seasons just less than a fifth of all households in both the full and subsample participated in tobacco production. Within the full sample of just under 12,000 households, only 1,588 indicated that they grew tobacco during the last cropping season. The subsample indicates a fairly similar proportion of engagement in tobacco production at 18.04 percent of households within this sample

[73] This question was only answered if they had indicated that they had grown tobacco in the last five cropping seasons. 1,551 households in the sub sample indicated that they had grown tobacco.
[74] To give an indication of the MK value in UK Sterling, 195MK was approximately £1 in 2004.

Figure 3.3.4 Household net income from tobacco sales during the last cropping season (subset)



Source: IHS2

3.3.8 Household enterprise and other income sources

In addition to agriculture, just fewer than 600 households in the full sample participated in some form of non-agricultural enterprise. Enterprises were most commonly owned by either the household head followed by the partner of the household head and were exclusively family owned businesses in almost all cases.

In addition to income from enterprise, a very small number of households also received income from other sources such as savings and investments during the past year. On the whole, very few households received income from such sources. A discussion of the income sources used within *Chapter 4* is outlined in that chapter.

3.3.9 Subjective assessment of well-being

Household heads were also asked about how people perceived some of the services they received and aspects affecting their quality of life. The responses are not particularly encouraging and are presented in Table 3.3.15.

Table 3.3.15 Households' perception of living standards (percentage indicating level of adequacy) (subsample)

Adequacy level	Indicator for standard of living			
	Food	Housing	Health	Clothing
Less than adequate for needs	55.39	51.22	59.22	70.76
Just adequate for needs	37.42	43.10	37.33	27.24
More than adequate for needs	7.18	5.68	3.45	2.00
Total	100.00	100.00	100.00	100.00
N	8593	8592	8591	8589
Missing	4	5	6	8

Source: IHS2

Fifty-five percent of household heads indicated that the food consumption within the household had been less than adequate for the household needs during the past month. Less than ten percent of households indicated that they had more than enough food for their needs.

Just over half of all households in the subsample also indicated that their housing was not adequate for their needs. Many individuals experience challenging living conditions, as was observed earlier. Therefore, it is not surprising that such a large percentage of households do not feel that they have appropriate housing conditions.

With respect to households' perception of the standard of health care that they receive, almost sixty percent of households indicate that the level of healthcare they have received is not adequate. It is difficult to be sure whether households are commenting on actual health services used, rather than, healthcare services that they would like to use. Therefore it is possible, that as a result of a lack of access or financial constraints, households deem that their standard of healthcare is inadequate. However, there is no way of knowing whether households have accurately interpreted this question as it was posed. Therefore, I would interpret this particular response with caution. The worst indicator is for clothing, with over seventy percent of households indicating that they do not have adequate clothing for their needs.

Households were then asked about where they viewed themselves on a series of steps. Step one being the lowest step, for the poorest, and step six is at the top, where the richest would stand. They were asked to place themselves on this set of steps, their neighbours and then their friends. Forty-five percent of households indicated that they were on the very bottom step; forty-two percent believed that their neighbour was on step two and 41 percent also believed that their friends were also on step two, which is indicative that they believe that their neighbours and friends are slightly wealthier than them.

Just over a fifth of households from the subsample indicated that their economic well-being had declined in the last year. Five percent indicated that they were significantly worse off in comparison to a year ago, whereas 56 percent indicated that they had been able to maintain or improve their standard of living over the last year. Thoughts about the future are generally positive, with just over two-fifths of households believing that their well-being would improve in the next year and 10 percent indicated that they believed they would be much better off a year on from the IHS2 as shown in Table 3.3.16.

3.3.10 Recent shocks to the household

In 2004, a new module was added to the IHS to gather information about shocks that a household may have experienced in the last five years and which affected household welfare. In the case of the IHS only negative shocks are taken into consideration, whereas in the KIDS dataset discussed later in this chapter households are asked about positive shocks.

Table 3.3.16 Households' perception of well-being (percentage indicating level of well-being)(subsample)

Level of well-being	Perception in comparison to a year ago	Perception in comparison to a year from now
Much better	3.05	10.37
Better	22.69	31.17
No change	30.80	36.90
Worse off	38.11	14.59
Much worse off	5.35	6.97
Total	100.00	100.00
N	8594	8589
Missing	3	8

Source: IHS2

Table 3.3.17 Recent shocks to the household (subsample)

Type of shock	Mean	Standard deviation	N	Missing
Lower crop yields due to drought or floods	0.629	0.483	8595	2
Crop disease or crop pests	0.244	0.430	8595	2
Livestock died or were stolen	0.353	0.478	8595	2
Household-business failure, non-agricultural	0.234	0.423	8596	1
Loss of salaried employment or non-payment of salary	0.107	0.309	8595	2
End of regular assistance, aid or remittances from outside household	0.080	0.080	8595	2
Large fall in sale prices for crops	0.378	0.485	8595	2
Large rises in price of food	0.780	0.414	8596	1
Illness or accident of household member	0.487	0.500	8595	2
Birth in the household	0.141	0.348	8596	1
Death of household head	0.043	0.202	8596	1
Death of working member of household	0.087	0.282	8596	1
Death of other family member	0.421	0.494	8596	1
Break-up of the household	0.101	0.302	8595	2
Theft	0.209	0.407	8595	1
Dwelling damaged, destroyed	0.106	0.308	8594	3

Source: IHS2

The type of shocks households are asked about fall into four main categories: agriculture, income, household members and dwelling shocks. The types of shocks and the percentage of households who experienced the shock are presented in Table 3.3.17. Almost two-thirds of all households in the subsample experienced poor crop yields in the last five years and 78 percent experienced large rises in food prices. Declining yields coupled with rising prices could substantially limit households' ability to access food which is a fundamental human need and may be indicative of a potential food insecurity problem, given that we also observed in the previous section that many households do not have adequate access to food.

Morbidity is a challenge within Malawi, with people experiencing both short term bouts of illness and longer term chronic health conditions. Therefore, it is not unexpected that nearly half of all households in the subsample indicate that illness or an accident of a household member has negatively impacted household welfare.

Deaths are common, with two-fifths of households in the subsample experiencing the death of a family member in the last five years. A much smaller proportion has experienced the death of working-age members which could have severe consequences for a household. Poor weather conditions which lead to drought and flooding clearly do not affect only individual households; they may also be much more widespread and affected large portions of a community, which puts a potential strain on support within communities in times of hardship. In contrast, the impact of a death of an individual tended to be predominantly concentrated within the immediate household and a smaller number of connected households.

The above discussion only provided a broad overview of the very large dataset, IHS2. Additional detail will be presented within *Chapters 4* and *5* where appropriate. The discussion will now move onto a review of the KIDS data that is subsequently utilised in *Chapter 6.*

3.4 Background to South Africa and the KwaZulu-Natal Province

This section presents a brief background picture of South Africa, including a short overview of the agricultural sector and the health challenges which the country experiences.

South Africa and the post-apartheid era

For almost fifty years South Africans experienced racial segregation under the apartheid regime which ended following election of the African National Congress (ANC) in 1994 and the election of Nelson Mandela as president. In the four elections that followed since 1994, the ANC continued their leadership within the country. The World Bank indicates that the development challenges within South Africa are linked to the legacy of the apartheid regime and there are two particular issues which South Africans face: persistent and extreme inequality along with inadequate education and health provisions for the very poor.[75] On the one hand, South Africa's economy has performed well according to the World Bank: inflation has been within target, annual GDP growth has been modest but steady at 3.3 percent per year[76] and there has been good overall debt and budget management. Moreover, the World Bank indicates that South Africa also benefits from well-developed infrastructure and institutions which help to foster an environment in which an economy can prosper.

Challenges within South Africa

Despite reasonable prosperity in the country, the World Bank indicates that a large number of people cannot gain employment. They indicate that it is this high level of unemployment which has contributed to the inequality which prevails within the country. The most recent estimate from 2009 indicates that 24 percent of the total labour force is unemployed and 14 percent are long term unemployed (World Bank, 2012).

Health and education are also significant stumbling blocks in terms of equality. Whilst there have been substantial improvements in access to health services, with 91 percent of the population having access to an improved water source in 2010 compared to 83 percent in 1994 and access to improved sanitation facilities has also increased over the same period from 72 percent to 79 percent (World Bank, 2012), the outcomes have not been felt by all.

As Figure 3.4.1 shows, the HIV prevalence rate in South Africa is high relative to some (although not all) of its neighbours. The high prevalence rate of 18 percent[77] places a severe burden on the healthcare system. Whilst the rate of HIV prevalence appears to have stabilised over the last decade, it had previously been rising quite significantly year on year as shown in Figure 3.4.1. A recent update on the financial implications associated with HIV/AIDS indicated that South Africa has been very successful in

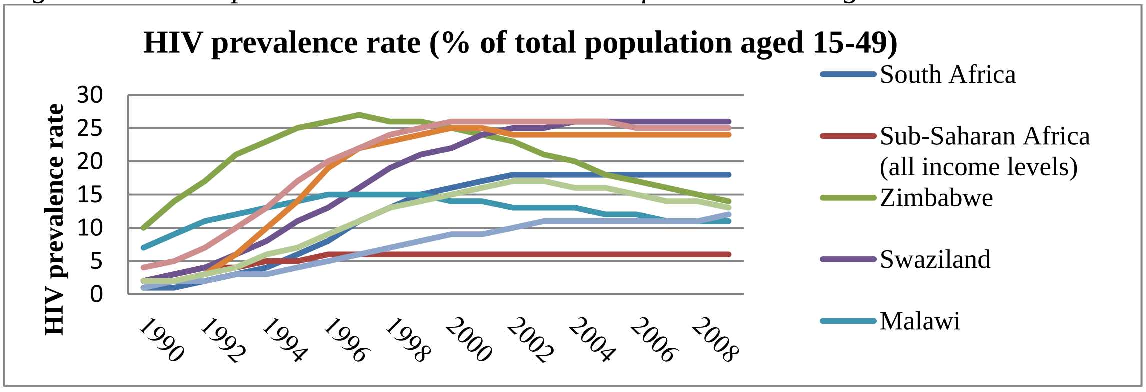
[75] http://www.worldbank.org/en/country/southafrica/overview
[76] http://www.worldbank.org/en/country/southafrica/overview
[77] 2009 figure, prevalence of HIV, total % of population aged 15-49.

responding to the HIV crisis with a particular emphasis on scaling up anti-retroviral (ART) care from 2003 onwards.[78]

In 2004, only three percent of the population with advanced stage HIV infection received ART and this has been steadily increasing to 37 percent based on the most recently available estimate from 2009 (World Bank, 2012). The full effects of the healthcare burden, associated with such a high rate of HIV infection, are yet to be realised and the country has committed itself to further increasing coverage of ART.

South Africa is not alone in facing a major challenge to finance ART therapy in order to help manage HIV. The World Bank concluded in early 2012 that significant investment needs to be made now in readiness for financing the ongoing need for treatment and even some of the richer African countries will face difficulties if such investment is not made.[79]

Figure 3.4.1 HIV prevalence rates within South Africa and its neighbours



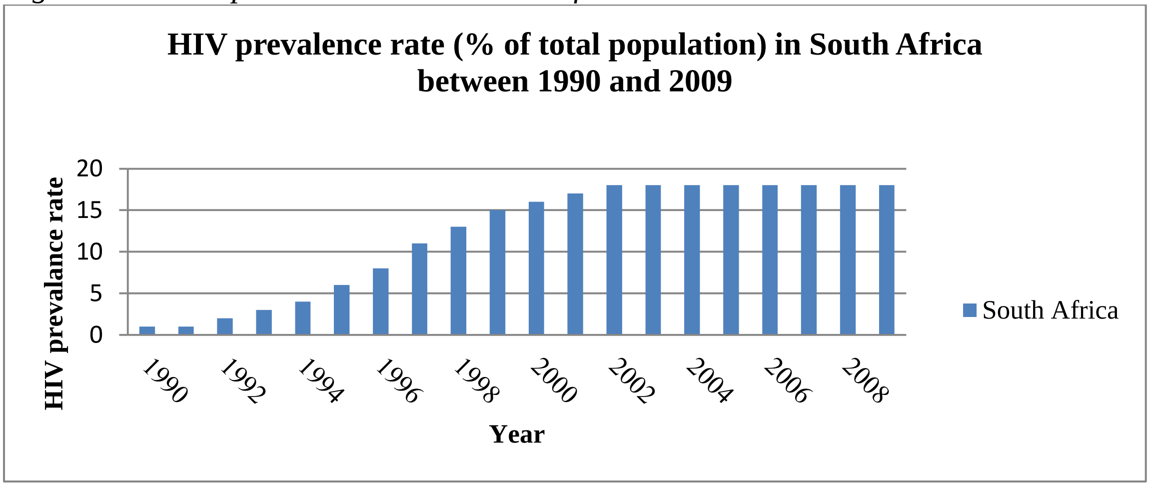
Source: World Bank (2012)

KwaZulu-Natal province geography

Just over a fifth of the population of South Africa reside in the KwaZulu-Natal province.[80] It is located on the eastern coast of South Africa bordering Swaziland and Mozambique. KwaZulu-Natal has one of the fastest growing urban centres in the world in Durban, which also hosts a busy shipping port.[81]

[78] http://blogs.worldbank.org/health/transitions-in-financing-hivaids-programs
[79] http://blogs.worldbank.org/health/transitions-in-financing-hivaids-programs
[80] http://www.southafrica.info/about/geography/kwazulu-natal.htm
[81] http://www.kznonline.gov.za/index.php?option=com_content&view=article&id=82&Itemid=174

Figure 3.4.2 HIV prevalence rate in South Africa between 1990 and 2009



Source: World Bank (2012)

To the west of the province is the Drakensberg mountain range, coastal areas are to the east with forest and grassland areas in between. The coastal areas are popular with tourists and for the cultivation of sugar and fruits whilst the middle area between the mountains and Indian Ocean is used for steel production and mining.[82] Together with the port in Durban, Richards Bay is also a very important area for KwaZulu-Natal, as it is the location of a number of large industrial firms engaged in mining and mineral production and processing. These firms together with other firms such as those engaged in car manufacturing have grown significantly and made a substantial contribution to foreign exchange earnings within the country.[83]

3.5 The KwaZulu-Natal Income Dynamics Study 1993-2004 (KIDS)

The KwaZulu-Natal Income Dynamics Study (KIDS) is a panel survey of a random selection of households in the KwaZulu-Natal province of South Africa with the initial wave being a subsample of the Project for Statistics on Living Standards and Development (PSLSD) which was one of the first nationally representative studies of poverty in South Africa (KIDS, 1998). During the apartheid era there was little interest in collecting information about poverty reduction so the PSLSD and the other surveys

[82] http://www.kznonline.gov.za/index.php?option=com_content&view=article&id=82&Itemid=174
[83] http://www.kznonline.gov.za/index.php?option=com_content&view=article&id=82&Itemid=174

that have followed are very important (KIDS, 1998). Although the PSLSD is useful in many respects, it is unable to address some issues which need to be examined in a dynamic setting. The creation of the KIDS survey sought to address this gap (KIDS, 1998).

The very first national household survey in South Africa took place in 1993 and was carried out by a collection of groups with experience in survey work and a number of universities led by the South African Labour and Development Research Unit (SALDRU) with support from the University of Cape Town, the World Bank and governments in Denmark, the Netherlands and Norway (KIDS, 1998). The PSLSD is very similar to the LSMS, which I outlined at the beginning of this chapter, in that it surveyed households on a range of topics including household demographics, education, expenditure and employment.

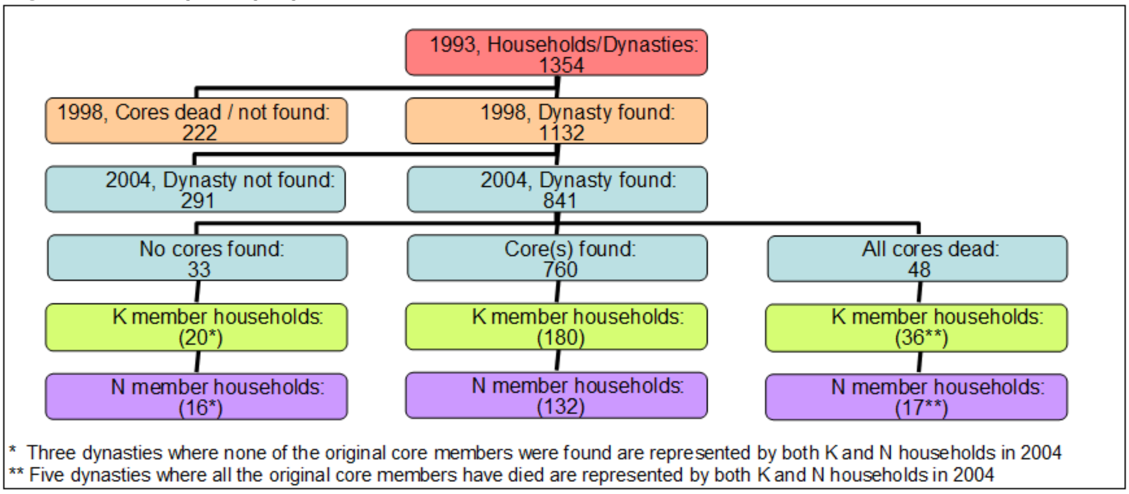
Following the initial PSLSD survey, households within the KwaZulu-Natal province were revisited and re-surveyed for what became the KIDS survey. The KIDS survey is comprised of three waves which were undertaken in 1993, 1998 and most recently in 2004. During the initial wave 1,354 households were interviewed. During the second wave in 1998 a total of 1,132 of these were re-interviewed, representing 84% of the original household dynasties and in the most recent wave 841 of the original households were found and re-interviewed.[84] The KIDS survey has been led by the University of Natal, the University of Wisconsin and the International Food Policy Research Institute (KIDS, 1998).

The selection of the KwaZulu-Natal province was partially determined by practical issues. These included the research interests of the team, resources available and the ability of the team to track down households who had been initially interviewed in 1993 (KIDS, 1998). Between 1993 and 1998 some slight changes were made to the survey; in particular inclusion of people not living in the household but who were linked to the household in an economic sense and the addition of four new modules. The new modules were economic shocks, social capital, assets from marriage and household

[84] In the 2004 wave, children who were included in the original core households but who are now adults themselves and have set up new households are included as they provide a tracking mechanism for the family dynasty

decision making (KIDS, 1998). In 2004, more modules were added. These included a module on Child Support Grants, recent deaths of household members, learning tests for children and work history of those aged 24 to 30 years old (KIDS, 2007).

Figure 3.5.1 Dynasty dynamics in KIDS: 1993-2004



Source: KIDS (2007)

As with the Malawi IHS2, survey enumerators underwent two weeks training prior to the survey being rolled out. Once the fieldwork began, the survey was conducted in parts, as it was very lengthy, so enumerators visited households on more than one occasion to complete the whole survey (KIDS, 1998). During the 1998 and 2004 surveys the definition of the household decision maker was redefined by the survey team. Previously a self-declared household head was taken to be the main decision maker. In reality it was felt by the researchers that this may not be the case and so key decision makers were renamed as "core" persons. A household member was designated as being a "core" person if they met one of the following criteria (KIDS, 1998):

1. They were a self-declared household head (from the 1993 survey).
2. They were the spouse/partner of the self-declared household head (from 1993).
3. Lived in a three generation household *and all of the following were true:*
    - Child, child-in-law, or niece/nephew of the self-declared head
    - At least 30 years old
    - Have at least one child living in the household
4. They were the spouse/partner of a person satisfying criterion (3).

The motivation behind this approach was to ensure that they captured more key decision-makers within the households by using a slightly broader definition than solely a self-declared household head. Considering core persons rather than a self-declared head also enabled the survey team to retain a greater number of households following households splitting into different households. This approach allows the researchers to create what they call a baseline of dynasties from the original 1993 survey.

In 2004, the study was adjusted again to account for the aging of the original 1993 core members and the prevalence of HIV/AIDS by including "children of the core or the next generation".[85] These are people who are now adults and have set up their own homes without their parents (KIDS, 2007). Moreover, in the 2004 wave, foster children were also included; these are children aged less than 18 years of age, who are associated with core and next generation households but haver no longer living with their own parents.[86] Figure 3.5.1 outlines the dynasty dynamics of KIDS from 1993-2004.

3.6 Introduction to the KIDS indicators

In this section I will provide a brief overview of some of the data modules within KIDS which I use within *Chapter 6*. There are a number of modules which I do not use. Therefore, I have not provided any additional detail on such modules. Some variables have been transformed for the purpose of the empirical analysis I have undertaken and where this is the case an explanation has been provided.

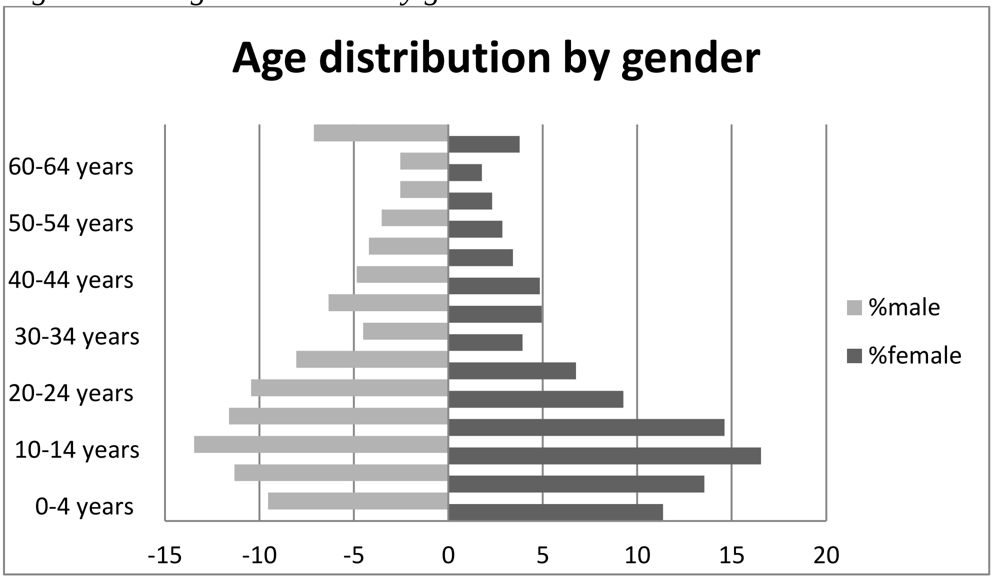
3.6.1 Household roster information

In all three of the KIDS surveys, information was collected on all the members of each of the households that were surveyed and tracked over the three waves. In order to be classified as a household member, the individual needed to have lived in the household for at least 15 days or more in the past year, share a common food source and share in/contribute to a common resource pool.

[85] These are known as "K".
[86] These are known as "N".

The discussion presented in each section will first outline the 1998 data and then the 2004 data, using the definition outlined above to ensure that the appropriate household members are included. Given that the full sample is not utilised in *Chapter 6,* the discussion presented here will also only consider the subsample for 2004. Moreover, given that the discussion is focused on the household head in *Chapter 6* and descriptive statistics associated with the sample (and therefore the household) I will provide some additional detail about the individuals within KIDS in this section where appropriate to give an overview of the data. There are no location indicators unlike in the IHS2 as the KIDS dataset is specific to one province within South Africa.

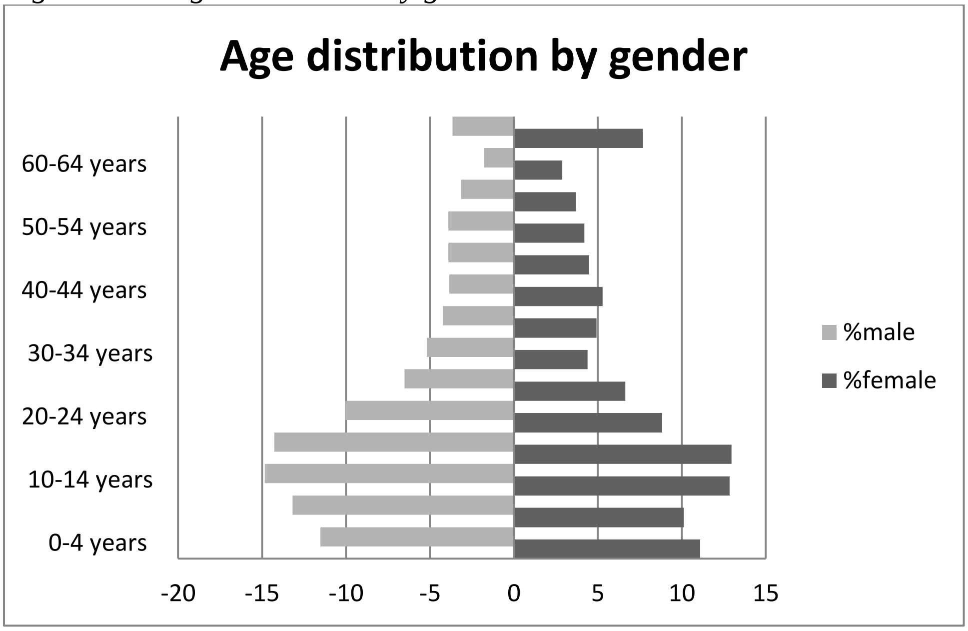
Figure 3.6.1 Age distribution by gender 1998



Source: KIDS (1998)

Figure 3.6.1, Figure 3.6.2 plus Table 3.6.1 provide an indication of the overall balance of individuals sampled within the KIDS surveys. The age profile is similar to that of Malawi; a larger proportion of young people and relatively fewer older households' members. A noticeable difference between 1998 and 2004 however is the switch and increase in the proportion of older women in the sample relative to older men.

Figure 3.6.2 Age distribution by gender 2004



Source: KIDS (2004)

Table 3.6.1 Distribution of gender by age

Age group	1998		2004	
	Gender		Gender	
	Male	Female	Male	Female
0-4	9.55	11.37	11.53	11.09
5-9	11.31	13.55	13.18	10.12
10-14	13.45	16.55	14.84	12.85
15-19	11.61	14.62	14.26	12.96
20-24	10.44	9.26	10.03	8.83
25-29	8.05	6.76	6.51	6.64
30-34	4.50	3.93	5.18	4.38
35-39	6.34	4.94	4.23	4.92
40-44	4.85	4.85	3.86	5.28
45-49	4.20	3.42	3.90	4.49
50-54	3.53	2.85	3.95	4.20
55-59	2.54	2.32	3.90	3.70
60-64	2.54	1.79	1.78	2.87
65+	7.19	3.78	3.65	7.68
Total	100.00	100.0	100.00	100.00

Source: KIDS (1998) and KIDS (2004)

In South Africa a child is not eligible to attend school at reception year level until the age of four. Therefore the discussion below excludes those household individuals who are aged less than five. The level of education achieved is reported as the highest level that the individual had at the time of the survey as a series of grades. Within South

Africa, grades one through to nine represent compulsory education or the General Education and Training. Hereafter there are a further three grades (10-12) which are classified as Further Education and Training. In order to then attend university grade 12 must be completed.[87] In the analysis that follows, education dummies were created where a value of one denotes that the individual possesses that level of education and a zero if they do not. Unlike the IHS2, education is well reported within the KIDS survey with only a small number of missing cases.

Table 3.6.2 Education of individuals

Variable	Mean	Standard deviation	N	Missing[88]
1998				
No primary	0.126	0.332	6723	10
Foundation	0.309	0.462	6723	10
Intermediate	0.228	0.420	6723	10
Senior	0.204	0.403	6723	10
FET	0.132	0.339	6723	10
2004				
No primary	0.091	0.288	4523	88
Foundation	0.266	0.441	4523	88
Intermediate	0.233	0.423	4523	88
Senior	0.216	0.411	4523	88
FET	0.194	0.396	4523	88
No primary				

Source: KIDS (1998) and KIDS (2004)

In contrast to the IHS2 data, a relatively large proportion of individuals have managed to achieve at least primary school education and education levels generally are higher than in Malawi. This follows from the fact that Malawi is classified as a low income country, whereas South Africa is an upper middle income country. The relatively small proportion of those not having currently achieved primary education is likely to be associated with the fact that such individuals were too young to have completed primary

[87] http://www.southafrica.info/services/education/edufacts.htm#02

[88] The total number of people aged over 4 and therefore potentially in education in 1998 was 6,733. In 2004 the relevant sample was 4,611 individuals. Therefore the missing values represent cases missing from the subsample.

education at the time of the survey given the relatively high levels of education generally reported.

Table 3.6.3 Main activity of individuals (1998)

Activity	Mean	Standard deviation	N	Missing
Regular employment	0.123	0.329	7499	0
Casual employment	0.029	0.168	7499	0
Self-employed	0.020	0.139	7499	0
Housewife/child-rearing	0.053	0.223	7499	0
Unemployed	0.166	0.372	7499	0
Formal education/school/university	0.404	0.491	7499	0
Creche	0.172	0.378	7499	0
Retired/Pensioner	0.029	0.168	7499	0
Disabled	0.004	0.060	7499	0

Source: KIDS (1998)

Table 3.6.4 Main activity of individuals (2004)

Activity	Mean	Standard deviation	N	Missing
Regular employment	0.075	0.265	5152	67
Casual employment	0.056	0.229	5152	67
Self-employed	0.026	0.160	5152	67
Housewife/child-rearing	0.035	0.183	5152	67
Unemployed	0.216	0.412	5152	67
Formal education/school/university	0.368	0.482	5152	67
Creche	0.039	0.193	5152	67
Retired/Pensioner	0.085	0.280	5152	67

Source: KIDS (2004)

The activity of the household head is considered within the analysis undertaken in *Chapter 6*, and therefore it is perhaps more appropriate to consider an overview of all household members more generally within this chapter to provide a general overview of the full sample.[89] Within *Chapter 6,* a small number of the activity categories are merged relative to those that are presented here. Specifically, those that indicated that they were a housewife or child-rearing, in a crèche or retired/pensioner were recoded with the use of a dummy and classify them as those not in the labour pool. A relatively

[89] In the case of the 2004 data, the full sample refers to individuals within the 829 households subsequently used within the analysis of chapter 6.

small proportion of all individuals in 1998 are engaged in employment activities, many are still in education or in a crèche (this is not unexpected though given the young population).

3.6.2 Household dwelling and services

Information regarding the type of home and facilities that household members have access to is utilised within *Chapter 6* as part of the participation model for the propensity score matching models and is therefore reported in full within *Chapter 6.* The only additional information to add here is a brief examination of home ownership and the size of the homes people live in. In 1998, eighty-four percent of households in were owned, with the remainder being rented. The proportion of home owners rises by a reasonable amount in 2004 to 92 percent.[90] The physical size of the home varies a great deal. Four rooms was the median number of rooms within the 1998 sample, rising to 5 in the 2004 sample.

3.6.3 Household assets

Each household was asked whether or not the household owned a series of durable assets,[91] who owned the asset and the approximate value of the asset. In both 1998 and 2004, an examination of the total value of assets across households indicates that there is a lot of variation. The variation in total asset value across households seems to vary quite substantially for two reasons: some households hold many assets whilst others hold few and the reported value of assets diverge by very large amounts.

[90] I also examine whether home ownership is sensitive to the choice of sample in 2004. The proportion of homeowners for the complete 2004 sample is 89 percent.

[91] The questionnaire indicates that respondents are asked about each asset and are able to indicate that they do own such an asset or not. However the dataset only reports the ownership of the asset if the response was "yes". Therefore all non-owners are coded as zero in Table 3.6.8.

Table 3.6.5 Durable asset ownership in KIDS 1998

Type of asset	Mean	Standard deviation	N	Missing
Radio/Hi-Fi stereo, TV/VCR/DVD	0.875	0.331	1212	0
Primus cooker, hot plate, electric or gas stove	0.869	0.338	1212	0
Fridge/freezer	0.527	0.500	1212	0
Electric kettle, toaster, other kitchen appliances	0.420	0.494	1212	0
Telephone/Mobile phone	0.119	0.324	1212	0
Household furnishings	0.905	0.293	1212	0
Bicycle	0.091	0.287	1212	0
Motor vehicle	0.156	0.363	1212	0
Jewellery	0.474	0.500	1212	0
Tractor or other farm vehicles	0.006	0.076	1212	0
Farm equipment (pump/plough)	0.129	0.335	1212	0
Other farming tools (spade/hoe)	0.485	0.500	1212	0

Source: KIDS (1998)

Table 3.6.6 Durable asset ownership in KIDS 2004

Type of asset	Mean	Standard deviation	N	Missing
Radio/Hi-Fi stereo, TV/VCR/DVD	1	0	829	0
Primus cooker, hot plate, electric or gas stove	1	0	829	0
Fridge/freezer	1	0	829	0
Electric kettle, toaster, other kitchen appliances	0.928	0.259	829	0
Telephone/Mobile phone	1	0	829	0
Household furnishings	1	0	829	0
Bicycle	0.189	0.391	829	0
Motor vehicle	0.367	0.482	829	0
Jewellery	0.614	0.487	829	0
Tractor or other farm vehicles	0.017	0.129	829	0
Farm equipment (pump/plough)	0.123	0.329	829	0
Other farming tools (spade/hoe)	0.971	0.168	829	0

Source: KIDS (2004)

3.6.4 Other sources of household income

The income measures used within *Chapter 6* are aggregates of some of the categories of income outlined in Table 3.6.7 and Table 3.6.8

Table 3.6.7 Sources of income 1998

Income source	Mean	Median	Min	Max	Standard deviation	N	Missing
Total income	2700.23	1474.17	-4459	75431.73	4300.21	1171	41
Casual wage	1347.08	238.33	0	18726.67	2173.18	1171	41
Other income non-labour	415.66	0	0	65470	2684.282	1171	41
Transport subsidy	12.85	0	0	4300	145.43	1171	41
Food subsidy	5.38	0	0	430	33.80	1171	41
Housing subsidy	42.09	0	0	2600	217.57	1171	41
Remittances	119.30	0	0	4000	216.32	1171	41
Net agriculture	49.80	0	-529.75	7357.50	317.82	1171	41
Self employment income/profit	221.0	0	-400	40000	1783.54	1171	41

Source: KIDS (1998)

Earned income is defined as the total monthly income[92] the household received minus subsidies, any income from non-labour activities and remittances. Unearned income is the sum of remittances, non-labour income and subsidies. The values reported are in South African Rand[93] and represent nominal monthly income from the various sources. These sources of income are classified as unearned sources in the subsequent analysis that is explored within *Chapter 6.* Income varies substantially across households in both 1998 and 2004. Although some of difference could be associated with reporting error, as income data in the absence of formal records may be unreliable. As the reported data is monthly, it is also possible that the reported figure could be affected by seasonal fluctuations in activity. An examination of the unearned income; income from subsidies and remittances shows that in general the amount of subsidy received by the average

[92] Total income is reported within KIDS and is inclusive of wage income, income from agriculture and self-employment.
[93] To give an indication of the Rand value in UK Sterling, 100 Rand was approximately £9 in 2004.

household is quite small, and many households receive nothing at all in both 1998 and 2004. Agricultural activities are calculated as a net income; the difference between revenue and cost, and therefore the negative values reflect cases where households expenditure has exceeded their revenue in that particular month. Self-employment income/profit may also be negative for the same reason.

Table 3.6.8 Sources of income 2004

Income source	Mean	Median	Min	Max	Standard deviation	N	Missing
Total income	4119.79	1903.58	-5983.33	97550	8366.43	817	0
Casual wage	242.39	0	0	7500	633.07	829	0
Other income non-labour	867.04	650	0	73800	3591.17	829	0
Transport subsidy	26.50	0	0	15400	563.28		0
Food subsidy	18.44	0	0	7700	312.02	829	0
Housing subsidy	35.81	0	0	2275	184.99	829	0
Remittances	127.00	0	0	3416.67	347.24	829	0
Net agriculture	73.32	0	0	13130.92	551.56	817	12
Self employment income/profit	96.51	0	0	15000	1084.65	829	0

Source: KIDS (2004)

3.6.5 Economic shocks

The shocks module within KIDS asks a household representative whether or not the household experienced a series of shocks, in the last 5 years in the case of the 1998 survey and in the last 6 years in the case of the 2004 survey. Such events can either be a negative shock or “bad surprise” for the household, or a positive shock or "good surprise". The types of shocks that are considered within KIDS are broadly similar to those considered within the IHS2 for Malawi, namely health- and agricultural production -related issues. Shocks are only reported if they occurred in the 1998 dataset and therefore it is assumed that in all other cases the shock did not occur. In 2004, households indicated whether the shock had occurred or not. The three most common types of shock are reported in Table 3.6.9 and 3.6.10.

Table 3.6.9 Recent shocks to the household (1998)

Type of shock	Mean	Standard deviation	N	Missing
Death of a household member	0.353	0.478	1212	41
Serious injury or illness	0.200	0.399	1212	41
Loss of a regular job	0.187	0.390	1212	41

Source: KIDS (1998)

Table 3.6.10 Recent shocks to the household (2004)

Type of shock	Mean	Standard deviation	N	Missing
Death of a household member	0.387	0.487	829	0
Serious injury or illness	0.231	0.422	817	12
Loss of a regular job	0.223	0.416	827	2

Source: KIDS (2004)

Positive shocks are not specifically considered within this thesis but are described here for completeness. The types of positive shocks considered within the survey are: new regular job for household member, new or increased remittances and/or government grant, inheritance/large gift/lottery winnings, big payment from a firm, scholarship for children or adults within the household. The two most commonly reported positive shocks in both 1998 and 2004 were securing a new job followed by a new or increased government grant. .It should be noted that government grants are inclusive of pensions, so it is reasonably likely such shocks could be associated with household members reaching pensionable age.

3.6.6 Health

The health module within KIDS is of interest as it contains information on a possible indicator of health shock that may be utilised within *Chapter 6*. As outlined in *Chapter 2*, one of the possible measures of a health shock is a decline in a self-reported health rating over time. The Grossman (1972) model predicts that the stock of health capital declines over time due to aging. Therefore a more substantial decline in health rating may be considered as being indicative of a health shock. Within KIDS current and previous health status is ranked on a five-point scale. This scale can therefore be used to potentially capture a health shock.

This measure of health is only proposed to core household members[94] and therefore is missing in a number of cases when considering the whole sample. This indicator has the possible advantage that it can capture health shocks for the individual and therefore enables me to explore whether the effect of a shock is sensitive to who the shock occurred to; a result that has been found in the existing literature. However, as this question is only posed to the core members, it ignores the vast majority of household members. Tables 3.6.11-3.6.14, present the information on health ranking for such core members.

Table 3.6.11 Subjective assessment current health status (1998)

Status	Current health status			
	Mean	Standard deviation	N	Missing
Excellent	0.134	0.341	1900	2
Very good	0.143	0.350	1900	2
Good	0.300	0.458	1900	2
Fair	0.289	0.453	1900	2
Poor	0.134	0.341	1900	2

Source: KIDS (1998)

Table 3.6.12 Subjective assessment of current health status relative to previous health status (1998)

Status	Current health status			
	Mean	Standard deviation	N	Missing
Much better	0.125	0.331	1900	2
Somewhat better	0.137	0.344	1900	2
About the same	0.446	0.497	1900	2
Somewhat worse	0.212	0.409	1900	2
Much worse	0.080	0.271	1900	2

Source: KIDS (1998)

[94] Core members are defined as an individual who was the household head in the previous wave of the survey, or because they meet all three parts of a criteria that identifies a core member. These criteria are: they were over 30 years of age in the previous wave, are the child or niece/nephew of the previous wave household head and are the parent of a child within the household roster.

Table 3.6.11 shows that less than one-fifth of core household members in 1998 indicated they were in poor health. Core members in 2004 have a similar experience, with less than ten percent indicating that they considered their health to be poor. Table 3.6.10 also indicates that health status has remained unchanged for almost half those surveyed in 1998 relative to the previous period. Around one fifth of the sample, however, indicated that their current health was somewhat worse than it had been in 1993 in the case of the 1998 data. The 2004 core members indicate a slightly different viewpoint; just fewer than one third indicate that their health is about the same in 2004 as it was in 1998 and two-fifths indicate that their health has declined.

Table 3.6.13 Subjective assessment current health status (2004)

Status	Current health status			
	Mean	Standard deviation	N	Missing
Very good	0.093	0.290	1191	0
Good	0.201	0.401	1191	0
Fair	0.278	0.448	1191	0
Poor	0.303	0.460	1191	0
Very poor	0.093	0.291	1191	0

Source: KIDS (2004)

Table 3.6.14 Subjective assessment of current health status relative to previous health status (2004)

Status	Current health status			
	Mean	Standard deviation	N	Missing
Much better	0.043	0.203	1191	0
Better	0.094	0.292	1191	0
About the same	0.311	0.463	1191	0
Somewhat worse	0.390	0.488	1191	0
Much worse	0.126	0.332	1191	0

Source: KIDS (2004)

3.7 Conclusion

I have presented a general overview of some of modules within the IHS2 and KIDS. There are of course additional questions which were asked of both individuals and household heads which I have not particularly touched on in any detail as they have not been used in the subsequent analysis that I have undertaken. The IHS2 includes additional modules that examine: security and safety, consumption of food,[95] non-food expenditure,[96] gifts received and given, social safety nets, credit and child anthropometry but such modules have not been considered within the analysis presented. Within KIDS some additional areas of questioning that I do not explore within *Chapter 6* are social capital and trust, and a module on children where the latter module includes child schooling, learning and anthropometry.

[95] Households are asked about reference periods of three days and also one week.
[96] Households are asked about reference periods of one week, one month, the past three months and the past 12 months.

Chapter 4: Childhood mortality in Malawi

Economics as a discipline enables us to bring together different fields within the social sciences such as demography, anthropology, politics and statistics, to collect and analyse data from the various agencies and organisations. Economic models provide frameworks from which we can develop an understanding of key relationships that exist between different variables and enable us to identify indicators which are most relevant when we are collecting data. Economic models enable us to use our knowledge about the complex relationships that exist between different factors to identify appropriate policy interventions. In so doing, we seek to have a better understanding of how such relationships differ in different contexts and identify and interpret the role of certain indicators to prioritise and inform decisions within particular areas of policy intervention.

The aim of this chapter is to use the framework developed by Mosley and Chen (1984) to analyse the determinants of child survival in Malawi. As outlined in *Chapter 2*, the framework was an important step forward as it brought together the disciplines of social and medical science to enhance our understanding of the factors which may influence survival. The framework specifically considers child survival within developing countries where social, economic, biological and environmental factors are shown to influence the probability of a child survival within their first five years of life. The model identifies what have become known as the proximate determinants: factors that directly influence the probability of survival and therefore morbidity and mortality. Social and economic determinants within this model operate through such proximate determinants and are grouped into three categories; individual, household and community level variables which will be explored within this chapter.

This chapter also takes into consideration the Grossman (1972) model of health. As discussed in *Chapter 2*, in this model, health is a capital good which depreciates over time. Therefore people undertake investment to maximise their stock of health given an initial endowment of health which is inherited. The Grossman (1972) model and the Mosley and Chen (1984) framework complement one another. The Mosley and Chen (1984) framework links together the role of the proximate determinants, whilst the Grossman (1972) model shows how health investments are determined by decisions made by the household for direct inputs such as the choice of *"medical care, diet, exercise, recreation and housing"*[97] (Grossman, 1972). Grossman (1972) also considers what he refers to as "environmental factors" which influence household production; a key factor is education, which is also highlighted within the Mosley and Chen (1984) framework.

Current evidence on the role of income within the literature has not been conclusive and given a lack of consensus therefore warrants further exploration. The argument put forward in *Chapter 2* by Case et al. (2002) is that wealthier families are more able to purchase investments in health which in turn raises the stock of health at any given point in time. Such purchases include items such as use of medical care, more nutritious food and safer living environments. As both the Grossman (1972) and Mosley and Chen (1984) frameworks demonstrate, these factors have a significant influence on the probability of survival.

Case et al (2002) find evidence to support the view that long-run household income has a positive effect on children's health status within the United States. Children living in households with higher incomes are more able to make investments in their health. They conclude that higher income households are as a result able to slow the rate of depreciation of health over time relative to those in poorer households. Ssewanyana and Younger (2005) also find evidence that increased household income proxied by an indicator for household wealth reduces infant mortality rates; although the overall effect is very small. Casterline et al. (1989) find a great deal of variation in their review of various studies which have included income or a proxy indicator of income. As noted in *Chapter 2*, in some cases the role of income is not explored due the lack of available data.

In contrast to the work conducted by Manda (1999), this chapter will use the IHS2 for Malawi in preference to the DHS given that the IHS2 does include some information on household income and this was not an indicator included within the Manda (1999) study. I believe that such a choice represents an opportunity to add to the existing

[97] p.225.

literature by exploring the potential role of household income in an examination of the determinants of childhood mortality within Malawi.

A further extension considered within this chapter is an exploration of the determinants of mortality by cause of death, illness vs. non-illness related deaths, to observe whether there are any important differences associated with the proximate determinants which may need to be taken into consideration.

This chapter will proceed as follows. *Section 4.1* will examine the patterns of childhood mortality across income groups with specific reference to Malawi. *Section 4.2* will consider the key causes of death in children across the globe and with reference to Malawi alongside a brief recap on the empirical literature. *Section 4.3* presents a discussion of the variables used to examine the determinants of child mortality in Malawi. In addition, some descriptive statistics of these identifiable key determinants are also presented along with an outline of the modelling approach used. *Section 4.4* presents the regression estimates and *section 4.5* concludes this chapter.

4.1 Patterns of child mortality and the MDGs

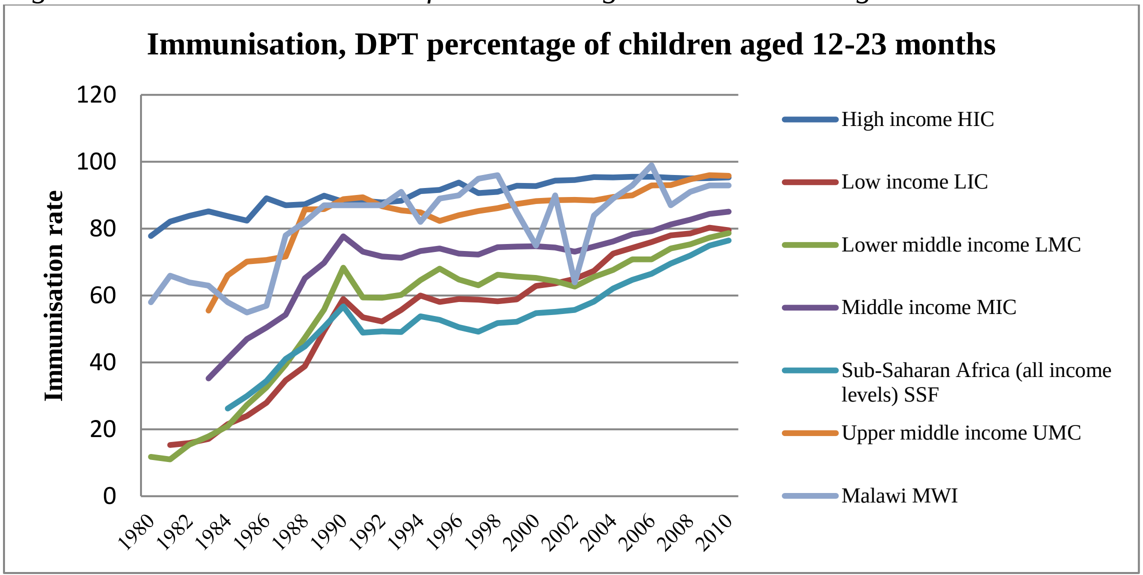
One of the key Millennium Development Goals (MDG) is to reduce child mortality by two-thirds by 2015. According to a World Bank update on the MDG, child deaths have been halved in recent decades.[98] This has been primarily attributed to improved health care, better nutrition and increased standards of living. As Figures 4.1.1 and 4.1.2 show, immunisation rates have been generally steadily increasing since the 1980s in all areas of the world. The one exception to this is Malawi, where immunisation rates against DPT dropped somewhat in the late 1990s. The World Bank[99] update indicates that over 20 years ago, 13 million children aged under five died in the developing countries from conditions such as diarrhoea, malnutrition, pneumonia, AIDS and malaria. By 2006, this figure had fallen to 10 million, which is a considerable improvement. What is most saddening is that many of these childhood deaths are preventable by ensuring that children have access to clean drinking water and are immunised against some of the

[98] http://www.worldbank.org/mdgs/child_mortality.html
[99] http://www.worldbank.org/mdgs/child_mortality.html

more common childhood diseases. There is still a considerable way to go. The World Bank indicates[100] that more than half of the MDG countries will not achieve their target of reducing childhood deaths by 2015[101] (see Figure 2.4.1). They note that reducing childhood mortality is particularly challenging in Sub-Saharan Africa due to conflict and HIV rates in the region.

To put the scale of the problem into context, Figure 4.1.3 shows the under-five childhood mortality rate by income level and in particular for Sub-Saharan Africa and Malawi. It is clear that childhood mortality rates have been falling in all cases since the 1960s, although strikingly there are much higher rates of childhood mortality experienced in Sub-Saharan Africa relative to low income countries as a whole since the early 1990s.

Figure 4.1.1 Immunisation rates for children aged 12-23 months against DPT[102]



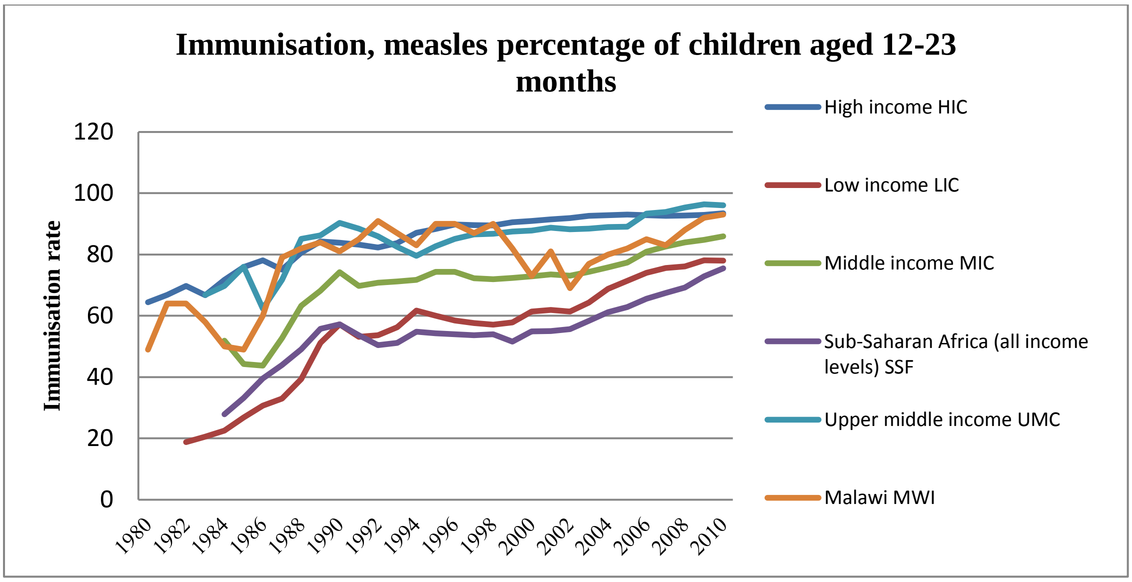
Source: World Bank (2012)

Moreover, what is very noticeable is the very high rate of childhood mortality experienced in Malawi in the 1960s. However, Malawi appears to have made some significant progress, as the rate of childhood mortality has been steadily falling over the

[100] http://www.worldbank.org/mdgs/child_mortality.html
[101] In 2008, it was reported that 107 out of 145 will not achieve this particular MDG target by 2015.
[102] DPT are the diseases diphtheria, polio and tetanus.

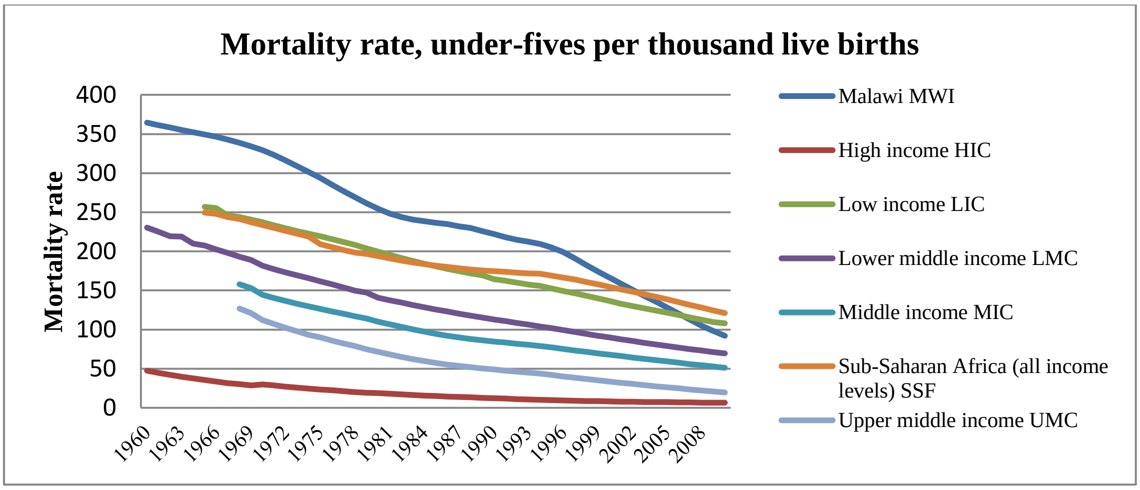
last five decades and is now lower than the rate reported for Sub-Saharan Africa as a whole.

Figure 4.1.2 Immunisation rates for children aged 12-23 months against measles



Source: World Bank (2012)

Figure 4.1.3 Childhood mortality rates by income level since 1960-2010



Source: World Bank (2012)

In this chapter, I will examine two indicators of child health, under-five and infant mortality in Malawi. Infant mortality refers to children who pass away before their first birthday; under-fives also include infant deaths. The reason for examining child mortality in this way is that the factors influencing under-five and infant mortality may differ. I will investigate the determinants of these two health outcomes using the Second Integrated Household Survey (IHS2) which is part of the Living Standards Measurement Survey.

As discussed in *Chapter 3*, the survey examines key socio-economic characteristics of households in Malawi including demographic, education, health, agriculture, labour-force and anthropometric indicators. I will now discuss child health and the factors that the literature outlined in *Chapter 2* indicates are commonly observed proximate determinants of childhood mortality.

4.2 Childhood health and mortality

In the previous section, I observed that estimates from the World Bank indicate that approximately 10 million children aged less than five die every year and in many cases it is asserted that these deaths would be preventable. The fact that these deaths are in many cases preventable has led to a great deal of research examining the causes and patterns of childhood mortality across the globe.

Figure 4.2.1 shows the most common causes of childhood deaths across the world and the leading causes of child death in Malawi as a comparison. There are some notable differences worth touching on briefly. One of the most immediately noticeable differences is the proportion of children who die due to HIV/AIDS related causes: 13 percent in the case of Malawi compared to 2% worldwide. Measles, malaria and pneumonia are also more common causes of death for children aged under five in Malawi compared to the rest of the world. Pneumonia is the leading cause of death in under-fives in Malawi and kills 1.6 million children every year across the globe (WHO, 2008). Worldwide the WHO statistics show that pneumonia kills more children than AIDS, malaria and tuberculosis combined.

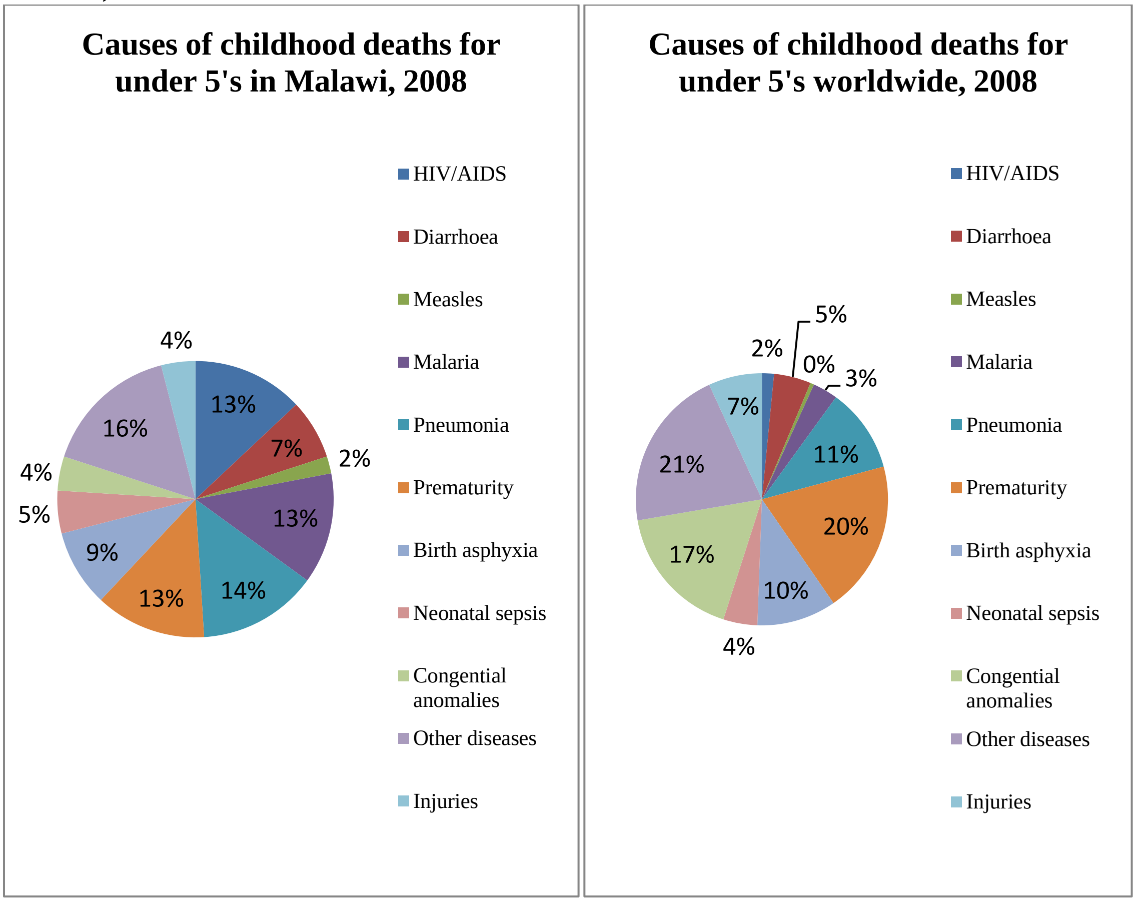
The WHO[103] state that the risk of death is greatest within the first month following birth and there are some key factors which have a strong influence on the likelihood of a very young baby dying. According to the WHO, the main risk factors to a new born are

[103] http://www.who.int/mediacentre/factsheets/fs178/en/index.html

premature birth, asphyxia and infection. The WHO indicate that these risks can be reduced by ensuring that women have access to (and use) good quality prenatal care, the baby is delivered by someone appropriately trained and high quality neonatal care.

This includes staff ensuring that the baby is kept warm following birth and given assistance to breathe if necessary. Once a new born is past the first month of birth, the risk factors that they will then face change slightly. The risk factors commonly experienced by under-fives are the contraction of disease, in particular pneumonia, diarrhoea, malaria, measles and HIV.

Figure 4.2.1 Leading causes of death in infants and children under 5 worldwide and in Malawi, 2008



Source: WHO Global Health Observatory Data Repository

Prevention is far more desirable than treating symptoms. When the immune system is weakened, a child's ability to fight everyday infections is compromised and hence their risk of contracting a secondary ailment increases. Exposure to secondary ailments will further weaken an immune system and further compromise the ability of the child to fight the condition and may lead to the additional risk exposure to other diseases and infections.

Therefore, prevention takes the form of providing an environment in which a child's immune system can thrive. The WHO[104] indicates that in over a third of cases, the death of a child can be linked to malnutrition and therefore this is a primary reason for advocating exclusive breastfeeding for the first six months. Exclusive breastfeeding[105] coupled with immunisation, access to clean water, and adequate sanitation play a significant role in helping to reduce the risk of contracting many conditions such as measles, pneumonia and childhood diarrhoea; as can the use of bed nets in the fight against malaria.

4.2.1 Approaches to examining child mortality

A key piece of literature surrounding the determinants of child mortality is that of Mosley and Chen (1984), who published a framework for studying child survival in developing countries. A more detailed discussion of the Mosley and Chen (1984) model can be reviewed in *Chapter 2*.

Empirical evidence presented in *Chapter 2* shows that the determinants of child mortality vary a great deal across countries. Houweling et al. (2006) find that the highest levels of child mortality in Kenya occur in areas where the mother was relatively uneducated and living in the rural area. In contrast, in Cameroon they find that the greatest levels of child mortality are observed amongst those where the mother was more highly educated. This could be because women who are more educated may also be employed and therefore spend less time breastfeeding/directly caring for their child. In Zimbabwe the greatest increase in child mortality has been in areas where the mother is highly educated and in urban areas (Houweling et al., 2006). Such differences therefore warrant further investigation.

[104] http://www.who.int/mediacentre/factsheets/fs178/en/index.html
[105] Assuming that the mother has adequate nutrition.

Furthermore, the literature surrounding the relationship between income[106] and child mortality also produces very mixed results. Anayamele (2011) finds that for six out of eleven countries in sub-Saharan Africa child mortality is negatively associated with wealth. Ssewanyana and Younger (2005) also find a negative relationship between infant mortality rates and wealth in Uganda. In contrast, income is found to have no effect on infant mortality in Egypt but large and significant effects on child mortality (Casterline et al., 1989). The discussion in *Chapter 2* indicates that income may influence health via its influence on households' ability to purchase health inputs. Poorer households who are constrained by income will be less able to make such investments in items such as medical care and nutrition which translates into children within those households having a lower stock of health.

Given that my study will examine Malawi, it is useful to briefly review the findings of Manda (1999). He notes that whilst Malawi has one of the highest rates of child and infant mortality there is relatively little research on the country. Manda (1999) specifically considers the relative importance of birth interval, maternal age at birth and birth order as determinants of childhood mortality in Malawi using data from the 1992 demographic health survey (DHS),[107] and a proportional hazards model. The survey sampled 4,878 women between the ages of 15-49 years and recorded all previous births. Manda (1999) however limits his analysis to births that occurred between January 1987 to September 1992. It is noted that the other indicators reported relate only to that period. Therefore such indicators are associated with current determinants of childhood mortality and not births which occurred prior to 1987.

In considering the data that I will be using, it is important to note that whilst indicators such as the level of education a mother possesses are unlikely to change over a relatively short period, other factors such as household income are very sensitive to the time at which the survey was taken. An event such as child mortality is highly likely to be influenced by both current and past decisions made by the mother/household. Capturing this dynamic in practice requires the use of panel data. The IHS2 is not such a

[106] Or an alternative measure of wealth which is used by some researchers.
[107] The first DHS survey undertaken in Malawi was in 1992, followed by surveys in 1996, 2000, 2004 and most recently 2010.

data set and, therefore, this does present a limitation to the study. This will be discussed in more detail later on within this chapter.

Additional “cleaning” of the Manda (1999) data also took place to remove cases where the date of birth and death could not be identified. I too have conducted a similar exercise to verify the number of children within my sample as accurately as possible. Manda (1999) finds that birth intervals, maternal education and age are statistically significant factors in explaining infant mortality: children of mothers with more education were less likely to die in infancy than those born to mothers with relatively little education. He also finds evidence that water source and types of toilet facility also have the expected (negative) effect on mortality, although the effects are not statistically significant. What is clear from the literature in *Chapter 2* is that there is no definitive evidence that certain factors influence infant and child mortality in *exactly* the same way across countries. This therefore, creates a case for continued exploration of this topic. As noted by Houweling et al. (2006) the differing findings reported in the body of literature suggest that there is an increased need to further examine child mortality at country level as adding to our knowledge and understanding can only be viewed as a positive outcome.

4.3 Data and variables

In order to reduce child mortality we need to be clear about the factors which pose the greatest risk and areas of opportunity in terms of policy intervention in different countries. Houweling and Kunst (2009) note that evaluating child mortality has substantial data requirements. This includes the registration of births, deaths and other relevant information. In practice, this can be rather limited in low and middle income countries. Consequently a common approach is to use household surveys and other types of large survey which collect information about a population. I have chosen to use the IHS2 as it contains a wealth of information about individuals and households within Malawi including information on health and importantly for my purposes household income which are the two key themes within this thesis. An alternative survey which one could consider is the Demographic Health Survey (DHS). The DHS is also a nationally representative household survey that collects information about indicators which can be used to monitor key areas such as population, health and nutrition.[108]

As previously noted in *Chapter 2,* the DHS surveys are one of the key datasets which are used for examining child mortality rates. The DHS surveys also include indicators on aspects of child health which I would like to explore, such as rates of breastfeeding, birth order and birth spacing which are identified as proximate determinants of child mortality and also considered by Manda (1999). However, the DHS does not include information on household income or expenditure (Devereux et al, 2006). Given the study by Manda (1999) used the DHS, I am proceeding with the IHS2 as an alternative, to build on the existing work that has been conducted within Malawi. The opportunity cost of using the IHS2 is therefore a sacrifice of not being able to examine certain indicators which I believe are important, at the gain of being able to consider income as an indicator.

4.3.1 Sample selection and dependent variable

As outlined in *Chapter 3*, 11,280 households were surveyed in the 2004 IHS2. For the purposes of my analysis I will only be considering a sub-sample: households where at least one child was born in the last two years and where there is only one potential mother in the household. This gives me a sample of 3,771 households and by definition 3,771 women as I will go on to explain. In order to identify my sample, I first use the information from the health module to identify those households where a birth has taken place. Women aged 12-49 years are asked whether they gave birth during the last 24 months (including all births which resulted in a still birth[109]): *"In the past 24 months, did you give birth to a child, even if born dead?"*[110].

This particular question enables me to clearly identify the mother in each case. This is important as I subsequently consider their characteristics; such as education, employment status, details about the birth and subsequent aftercare for the most recently

[108] http://www.measuredhs.com/What-We-Do/Survey-Types/DHS.cfm

[109] Ideally live births would have been sought as this is the traditional measure used. In the "deaths in household" module, stillbirth is not included as a cause of death. Therefore, I am unable to adjust the data to account for stillbirths to provide a figure for live births which would be my preferred indicator.

[110] Question D35 of the health module questions within IHS2.

born child. In addition, I only select those cases where there is only one mother per household. The reason for only selecting households where there is one mother is due to how deaths and the number of children within the household are reported. Deaths and the number of children are reported per household and not per mother. This creates an identification problem, which I will now outline with an example.

There are two (or more) women of child bearing age within a household; both women have had a child in the last two years. I therefore initially identify them as being relevant for my sample. One of these children subsequently dies and this is registered in the death module of the IHS2. The problem I have is that I cannot attribute this child’s death to one particular mother; I can only attribute it to the household.

Moreover, I also wish to consider the number of births that each mother has had within the last 2 years in my analysis as a proxy for birth spacing. I can identify all living children in the household; but I cannot attribute these living children to any particular mother. The household roster links each individual to the household head and not by their relationships with one another. Therefore in calculating an indicator for the number of births, I need to be clear about how many births are attributable to each mother. The only way to do this is to exclude all cases where there are multiple mothers within a household who were initially identified as having given birth within the last two years.

Ideally I would like to have complete and detailed information on *all* children who have been born in the last five years but this is not possible due to limitations of the data collected within the IHS2. Therefore, my choice of sample is selective as it does not include all mothers and by default not all children.

The identification of infant births and mortalities is less problematic as the sample is only reduced slightly in cases where there are multiple women within a household. All infants are included within the sample as by default they are less than two years old. Therefore both their birth and death are registered. For child mortalities the issue is very slightly different. Some households are excluded if a child has not been born recently and then just as with the case of infants further exclusions are made where there are multiple mothers. Such an approach is not ideal and *may* potentially lead to the introduction of selection bias; women may have had children more recently due to losses whilst others who have not experienced losses may have chosen not to have children more recently. Hence the sample may then represent particular types of women and their children. There is no way to circumvent this problem given the way in which the IHS2 data is presented.

Another reason for selecting only those cases where a mother has given birth in the last two years is that I cannot identify *any* characteristics surrounding the birth of a child if a birth did not take place within that period.[111] A mother may of course have had very different experiences delivering each of her children (in cases of multiple births within the last two years) which I cannot control for. Therefore, such indicators may not be wholly relevant where the deceased was an older child and/or in the small number of cases where a woman had more than one child in the two year period. A further concern is that I can only identify those children who are currently living or those who have died within the last two years. I have no way of identifying any children that were deceased more than 2 years ago.

Such data restrictions are far from ideal and would be avoided with the use of the DHS. However, I would then be unable to explore the inclusion of income and move beyond the work of Manda (1999). As a result, despite not being ideal, I have restricted my sample to only those households where a child has been born in the last two years *and* where there is only one mother per household.

Once these mothers (and by default their households) are identified, I then further verify the information with other information provided in the household roster, which provides details on all living children who given my previous sample selection can only have been to one particular mother within each household.[112] Within the deaths module,

[111] The details surrounding births only pertain to *the last born child* and not all births that took place within the last two tears for each mother.
[112] One potential issue is that a child may be adopted and this would cause me to introduce upward bias in the number of births that each mother has had. However; I have no way of separating adopted children from biological children due to the way household members are coded. .

some additional detail about the deceased such as their age, gender and cause[113] of death is also reported. One of the important issues to note is that the validity of the data on births and deaths is potentially subject to reporting error as the information provided by the household is not confirmed by any official birth and death records.

I note that there were a small number of inconsistencies between the number of children born and who had subsequently died and the number of surviving children reported within the household roster. I therefore cross referenced the different sets of information: the number of births, the number of living children[114] and the number of deceased children for accuracy. Where I could not reliably match the current household roster to the information reported on births and deaths, the case was excluded from the subsequent analysis.

The household roster represents *current* living household members, whereas birth and death information is based on recall and therefore may not be as accurate. Ideally, formal birth and death records would be preferable, but as noted by Houweling and Kunst (2009) very often this information is not readily available. The DHS also uses self-reported information on these areas. Therefore I have no reason to believe that the reporting error which may be present in the IHS2 is any more problematic than it would be with another data set such as the DHS.

I consider mortalities which have arisen in the last two years as this is the only data that is available from the IHS2: *"Over the* **past two years**, *did any member of your household die, including any infants?"*

Ideally, I would want to know about mortalities which have taken place over a longer period of time. One of the new modules "recent shocks to the household" within IHS2 considers whether the household has experienced a death in the last 5 years. However, there is no way of attributing this to a child. Therefore the only data which is available on deaths for children is those deaths that occurred in the last two years. Consequently,

[113] Cause of death can be attributed to old age, illness or other causes but this is at the discretion of the household head when they provide their answer and not confirmed by any medical professional.
[114] Reported within the household roster.

one of the limitations of my analysis is that I am unable to identify the death of any child under five who died more than two years ago.

Effectively my sample includes two groups: those that have experienced a mortality within the last two years and those that have not experienced child mortality in the last two years.[115] The DHS 2004 report for Malawi indicates that the under-five mortality rate was 133 deaths per 1,000 live births for the period 2000-4. This 13.3% mortality rate is similar to the calculated 14% mortality rate for my sample.

As outlined within *Chapter 2*, infants and children under five face differing risk factors which may affect the probability of mortality. Therefore, I examine infants and under-fives separately, to examine whether the determinants of mortality differ between infants and young children. Households report whether or not they experienced the death of a household member (including children) and the age of the deceased. I create a count variable for both infant and child mortality which will be used in my Poisson model. For each mother where no child or infant mortality occurred the count recorded is zero. Where a child or infant mortality does occur, the count recorded will be a positive integer. In my probit model, I create a dummy variable for infant and child mortality respectively where D=1 for a death and D=0 for no death. The information provided by each household within my sample indicates 579 children died in the two years prior to the survey and 373 of these deaths were infant deaths.

In order to apply the age cut offs as precisely as possible, I work with the age of the deceased in years and months. This enables me to ensure (although there may still be reporting error) that an infant is not treated as a child more generally and children over the age of five are not included in any count of an under-five mortality. An under-five death is therefore defined as the death of a child who does not reach their fifth birthday. This indicator includes all children aged up to and including 59 months of age.[116] Infants are classified as any child who does not reach their first birthday. Therefore, based on the information provided by the household, an infant mortality is inclusive of all children up to 11 months of age. For clarity, the dependent variable of under-fives is

[115] The comparison/control group is likely to include households which have experienced mortality more than 2 years ago and therefore it does not represent an ideal control group.
[116] As age is reported in months and years.

inclusive of infants. I employ as the dependent variable either a dummy and/or a count variable. The count variable relates to households which have experienced more than one death.

4.3.2 The presence and nature of childhood mortality in Malawi

As Table 4.3.1 shows, the majority of households (86% with children and 91% with infants) in the sample do not experience childhood mortality. For cases where a child does die, it is possible to examine the possible cause of death where reported by the household. Table 4.3.2 shows that in 64% of cases a child's death was reported as being illness-related.

Table 4.3.1 Number and frequency of under-five and infant deaths in the last two years (3,771 households)

Number of births	Frequency	Number of children under-five deceased	Frequency	Number of infants deceased	Frequency
0	-	0	3,246	0	3,427
1	3,450	1	475	1	318
2	302	2	46	2	23
3	19	3	4	3	3
Total births/deaths	*4,111*		*579*		*373*

Source: IHS2

Table 4.3.2 Cause of under-five and infant mortality for deaths occurring in the last two years

	Under-five		Infant	
Cause of death[117]	Frequency	Percent	Frequency	Percent
Illness	373	64.42	189	50.67
Non-illness	203	35.06	182	48.79

Source: IHS2

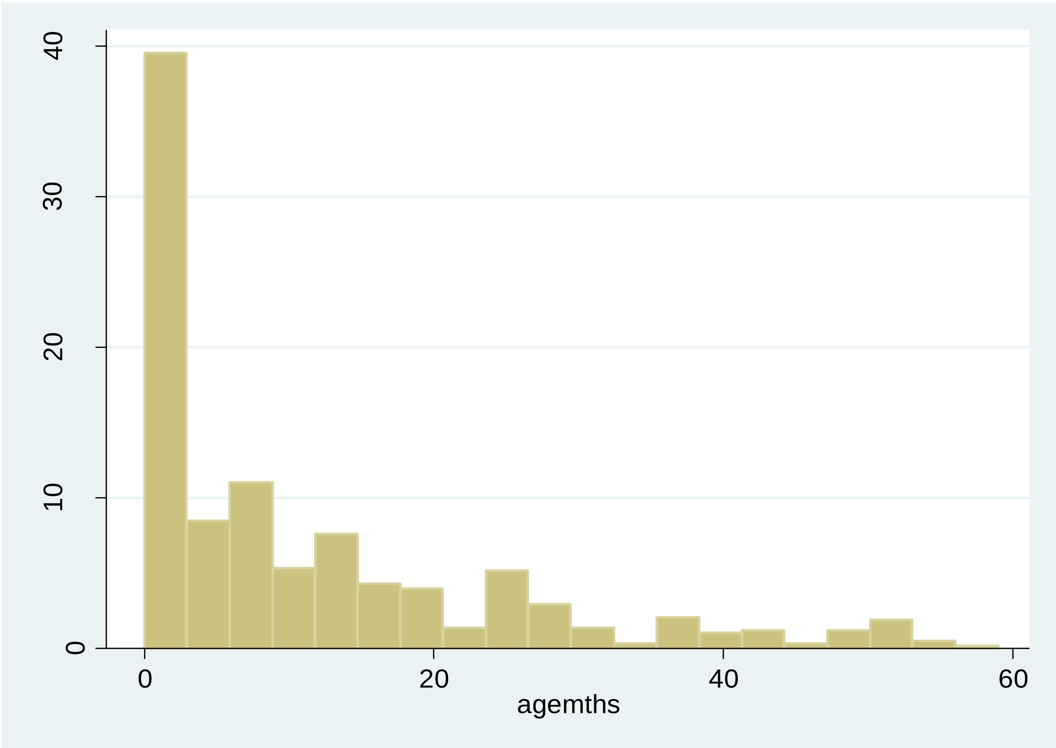
If we only consider infants (those aged less than one year), approximately 51% die from illness and 49% die from other causes[118]. Figure 4.3.1 below shows that the number of deaths related to illness declines as a child gets older. This result is not unexpected as a

[117] The cause of death was not reported in all cases. Hence, the number of deaths reported in Table 4.3.1, does not exactly correspond to those in Table 4.3.2.
[118] The two values reported in table 4.3.2 for both under-fives and infants do not add up to the total number of respective deaths as the cause of death was not reported in a small number of cases.

child's immune system would strengthen over time and this may have enabled them to fight infection for some time prior to death.

The majority of households reported how long a child was ill for before they passed away due to illness-related causes. The median number of days that a child was ill for before they died is reported to be 7 days. A quarter or more of children were ill for approximately a month before they passed away and one percent experienced illness for over a year before they died.

Figure 4.3.1 Age in months of under-fives dying from illness-related causes



Source: IHS2

Table 4.3.3[119] summarises the illness related causes of death reported by the household. Malaria, diarrhoea and pneumonia are the leading causes of illness related death for children in Malawi based on the information provided in the IHS2. In infants, the same pattern emerges where the leading cause of death is malaria responsible for 29% of infant illness related death cases.

[119] The illness cause of death was missing in 3 cases in relation to child deaths. Hence, the numbers of illness related deaths reported in Table 4.3.2 and Table 4.3.3 do not match.

Illness related causes of death have been confirmed by a medical professional in just over two-thirds of cases for under-fives. Whilst we may have some degree of confidence regarding the cause of death in these cases, it would be preferable to have the cause of death confirmed in a much larger proportion of cases.

Thirty-five percent of childhood deaths are attributed to non-illness related causes. The sources of non-illness related deaths are summarised in Table 4.3.4.[120] It is clear to see that birth complications are by far the most frequent non-illness related cause of infant deaths.. As we saw in section 4.2, infants are particularly vulnerable in the neonatal period. Ensuring that a child is properly warmed and aided to breathe after birth can make a significant difference to survival according to the WHO (2012). Therefore, it is not unsurprising that birth complications are the leading cause of non-illness related deaths.

Table 4.3.3 Illness related causes of child and infant mortality

Illness related cause of death	Child	Infant
Malaria	99	55
Measles	9	2
Diarrhoea	63	25
Pneumonia	57	34
Meningitis	7	2
Malnutrition	31	5
Tuberculosis	2	1
Heart disease	1	1
High blood pressure	8	3
Stroke	1	1
Cancer	1	0
Kidney disease	2	1
Liver disease	1	0
Diabetes complication	1	1
Unknown/refused/other	87	58
Total	*370*	*189*

Source: IHS2

In conclusion, it is clear that of the 579 children who died in the two years prior to the IHS2, a substantial proportion of these deaths in Malawi were attributable to illness. As Table 4.3.3 shows these deaths are most commonly related to malaria, pneumonia and

[120] The cause of non-illness related deaths is also not reported in a number of cases. Hence, the results presented in Table 4.3.4 do not match with those presented in Table 4.3.2 due to some missing data.

diarrhoea. At least a proportion of these deaths may have been avoidable by undertaking appropriate prevention measures, or, following illness, the use of treatment. Malnutrition is directly cited as the cause of death in only a small number of cases. However, this is not unusual in many respects as very often malnutrition is viewed as a secondary (underlying) factor to the cause of death.

Birthing complications cause more deaths in infants than malaria, pneumonia and diarrhoea combined. Therefore, it is useful to examine the type of health services a woman accessed during her pregnancy[121] in section 4.3.3 to see if there are any patterns that may emerge which could explain this result.

Table 4.3.4 Non-illness related causes of child and infant mortality

Non-illness related cause of death	Child	Infant
Traffic accident	2	2
Accident or injury	5	2
Birth complications	129	128
Murder	1	0
Witchcraft/sorcery	33	20
Other	30	27
Total	200	179

Source: IHS2

4.3.3 Maternal characteristics and pregnancy related health services

Age of mother and fertility

The information reported by each household provides some details about the characteristics of the mother and the health services she accessed during pregnancy. The age of the mother is reported for the time at which the survey was conducted. Only women aged 12-49 are asked whether they gave birth in the previous two years. Figure 4.3.2 shows the age of mothers in my sample. I note that there are a number of women at both extremities of the age spectrum.

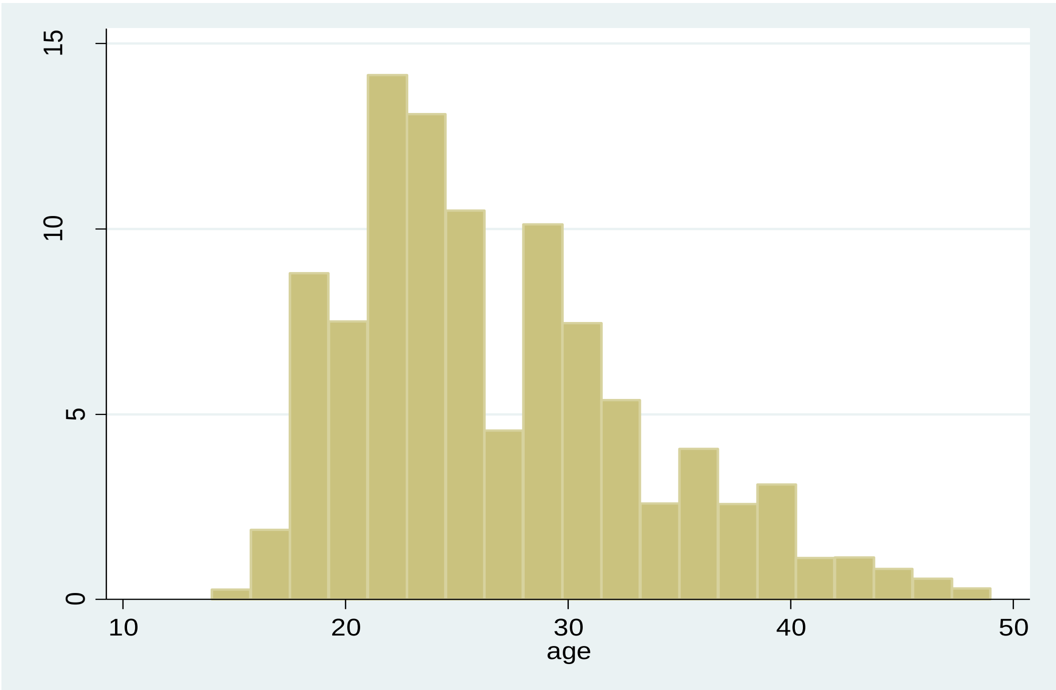
In *Chapter 2*, I noted that the age of the mother at the time of birth is likely to impact on the probability of child mortality, with higher rates of mortality to very young and older

[121] Responses relate to the most recent pregnancy and are therefore not necessarily reflective of behaviour in other previous pregnancies.

mothers compared to those in their 20-30s. To control for this a quadratic age term can be included as an independent variable.

It should be noted that as the mother has been asked about her current age rather than the age she was when she gave birth, age is accurate to within a 2 year period. Figure 4.3.2 shows that the majority of women give birth in their 20s and mid 30s which are the peak years of fertility. There are a number of women who are reported as nearing fifty years of age who have given birth in the last 2 years, according to the information provided. Further examination of the data, using the household roster and the individual identifiers suggests that there are some large reporting errors for age, which would explain the relatively large number of older mothers in this sample.

Figure 4.3.2 Present age of mother



Source: IHS2

It would be desirable to have additional information about each of the individual births each woman has experienced. This would enable me to examine birth spacing, which is found to have an impact on the rate of childhood mortality and is discussed in more detail in *Chapter 2*. However, this information is not provided by respondents in the survey. As a result, it is not possible to include birth spacing as an additional independent variable in the analysis. Birth spacing should, therefore, be considered as an omitted variable. I can, however, include a variable for the number of children born to each mother in the past 2 years.[122] This enables me to introduce *some* control for the fact that the mother may have had very closely spaced births. Whilst this is not an ideal indicator, it may provide some indication of recent fertility.

Fertility and child mortality are closely associated with one another (Rosenzweig and Shultz, 1983; Doepke, 2004; Ben-Porath, 1976). One of the reasons put forward to explain this relationship is what Rozenzweig and Shultz (1983) refer to as biological relationships. Women who have a young child that is still being breastfed are less likely to become pregnant due to what is known as lactational amenorrhea, a form of natural birth control;[123] the loss of such an infant may therefore reduce the gap between births. As a result of this, women who have recently lost an infant may be more likely to conceive.

A second issue posed by Rosenzweig and Shultz (1983) is that the probability of an infant death may be associated with the number of previous births and the mother's cumulative fertility, which in turn may affect the child's health. Clearly if mothers have a shorter gap between births, their body has less time to physically recover. As a result, this can have an impact on the health of any future children and therefore their probability of survival (Pebley and Strupp, 1987). One may therefore expect that, the more children that are born, the greater is the probability of mortality as the mother has experienced more physical strain on her body. In addition, closely spaced births may also be associated with increased competition between each of the children for limited resources as more children are born. Consequently, these factors combined would increase the probability of mortality. A third issue touched upon by Rosenzweig and Shultz (1983) which is a key theme raised within the child mortality and fertility literature is the concept of hoarding and replacement.

[122] As deaths are only reported for the last 2 years, I can only confirm births and deaths that have taken place in the last two years. Information from the household roster combined with data from the health and death module is used to calculate the number of births each mother has had in the last 2 years by only considering those children aged two and under. I note that within the household roster children refers both to biological children and adopted children. Therefore, there may be measurement error for the number of births if a household has adopted children.

[123] This is known as the lactational amenorrhea method (LAM). Women who are exclusively and continuously breastfeeding do not menstruate and therefore are regarded as infertile/sterile/less fertile in the postnatal period. As a result, they are far less likely to conceive and it can be argued that breastfeeding therefore acts a form of natural contraception.

Hoarding refers to the behaviour of parents who ex ante decide on the optimal number of children to have given their expectation of child mortality. Replacement is an ex post response, where parents have additional children in response to a child mortality (Ben-Porath, 1976; Benefo and Schultz, 1996; Rosenzweig and Schultz, 1983). One would expect that households which perceive that they are at a greater risk of child mortality ceteris paribus would, from a hoarding perspective, have more children than households who perceive that they are at lower risk of experiencing child mortality. Moreover, households which have experienced child mortality previously are also more likely to have additional children as they seek to replace the deceased children so that they can maintain their optimal household size. I therefore anticipate that the number of births experienced in the past two years will be positively associated with under-five (and infant) mortality.

Location of mother

Table 4.3.5 below shows the location of the mother and the corresponding number of births and deaths within each area. In total just over 4,000 children were born in the two years prior to the survey with the vast majority of these occurring in the rural area, although the birth rate (births per household) in the urban area is not substantially less than the rate observed in rural households. The number of births, child and infant mortalities per household are also reasonably similar across the three regions. The main striking difference is that the number of child and infant mortalities per household is lower in the urban area relative to the rural area. This is not unexpected however, as there may be better access to medical services within the urban area and women may also be more educated which would all contribute to a reduction in the risk of mortality.

Education of mother

Just under 80 percent of women in the sample who reported their level of education, report that they have no education at all and the birth rate is substantially higher for these women. The number of children born per mother[124] falls as the level of education rises.

[124] Due to the determination of the sample, a mother and a household are the same unit effectively.

This is likely due to the fact that more educated women may postpone having children and/or tend to have fewer children. Women who have more education may be better equipped to make decisions about controlling their fertility, although this is not universally agreed upon, (Cleland, 2002; Jeffery and Basu, 1996). The number of children born to educated women may also be fewer if there is an income effect. Women who are more educated may be able to command a greater wage in the market and hence additional children would represent an increase in the opportunity cost of having children.

Table 4.3.5 Number of children born and childhood and infant mortalities in the past 2 years by total, region, area and gender

		Number of children born	*Number of childhood mortalities*					
Overall			*All children*	*All infant*	*Female* children*	*Male* children*	*Female* infant*	*Male* infant*
	Total	*4,111*	*579*	*373*	*285*	*293*	*184*	*188*
	Per household	*1.09*	*0.15*	*0.09*	*0.08*	*0.08*	*0.05*	*0.05*
Region[125]								
North	Total	*604*	*63*	*43*	*29*	*34*	*19*	*24*
	Per household	*1.07*	*0.11*	*0.08*	*0.05*	*0.06*	*0.03*	*0.04*
Centre	Total	*1613*	*243*	*133*	*124*	*119*	*71*	*65*
	Per household	*1.10*	*0.17*	*0.09*	*0.08*	*0.08*	*0.05*	*0.04*
South	Total	*1894*	*273*	*197*	*132*	*140*	*94*	*102*
	Per household	*1.09*	*0.16*	*0.11*	*0.08*	*0.08*	*0.05*	*0.06*
Area								
Rural	Total	*3722*	*546*	*350*	*271*	*274*	*174*	*175*
	Per household	*1.09*	*0.16*	*0.10*	*0.08*	*0.08*	*0.05*	*0.05*
Urban	Total	*389*	*33*	*23*	*14*	*19*	*10*	*13*
	Per household	*1.07*	*0.09*	*0.06*	*0.04*	*0.05*	*0.03*	*0.03*

Source: IHS2 * Please note that one household in the rural south did not report the gender of the deceased child hence the values by gender differ from the total

Table 4.3.6 also shows that the number of child mortalities per household falls as education levels increase. As expected the number of child mortalities per household

[125] 564 households in the north, 1,468 in the centre and 1,739 in the south, 1 household did not report the gender of the deceased in the south and so this household drops out when looking at gender.

declines with increasing education as more educated women may possess more knowledge about disease; in particular its prevention and treatment (Caldwell and McDonald, 1982) and therefore would seek to minimise risk factors for their children so the opportunity for disease to take hold is reduced.

Use of health services

As noted in *Chapter 2*, the place and the person who assisted the delivery of a child can be extremely important, particularly in relation to infant mortality risk. In a clean environment, where trained medical staff are on hand, there should in principle (ceteris paribus) be fewer child mortalities. Each mother who had given birth in the last two years gave details about her use of health services during her most recent pregnancy and the subsequent birth.

Table 4.3.6 Number of children born and childhood mortalities in the past 2 years by education level of the mother

		Number of children born	*Number of childhood mortalities*	
Overall			*All children*	*All infant*
No primary	Total	*2318*	*350*	*238*
	Per household	*1.09*	*0.16*	*0.11*
Primary	Total	*292*	*20*	*14*
	Per household	*1.06*	*0.07*	*0.05*
Secondary	Total	*265*	*20*	*14*
	Per household	*1.04*	*0.08*	*0.06*
Tertiary	Total	*4*	*0*	*0*
	Per household	*1.00*	-	-

Source: IHS2 Note that a number of women did not report their level of education at all and are excluded from this Table (2,657 households included). There are 2,126 households where the mother has no primary education, 273 where she possesses primary education, 254 where she possesses secondary education and four cases where the mother has tertiary education.

The information presented in Table 4.3.1 shows that a number of women experienced more than one birth in the two years prior to the IHS2. In some cases this could be associated with multiple births, such as in the case of twins or triplets. However, given that the reference period is two years it is also equally possible that a woman could have experienced and therefore delivered two (or more) separate children, assuming a standard pregnancy of nine months. Ideally, I would like details for each individual birth but this is not possible. It is entirely possible that women who have given birth more than once have in fact used different health services for each pregnancy and birth which I am unable to capture.

Moreover, the use of health services may not be relevant for the examination of the mortality of children aged more than two years old. Once a child is beyond infancy, other environmental factors such as access to safe drinking water[126] are far more likely to have an influence on the likelihood of experiencing mortality; particularly in the case of illness related deaths. I anticipate the use of health services will reduce the number and probability of under-five and infant mortalities.

Table 4.3.2 shows that the most common, non-illness related cause of death in infants was birth complications. The expectation would be that a child born in a hospital or other similar medical facility would have a better chance of survival (all other things being equal) than a child born in a non-medical environment. However, there is also a counter argument to this; a more complicated and risky birth is more likely to take place in a medical facility (if the risk was known before labour started or indeed a mother could be moved to a facility during birth if there were signs of distress). Therefore, it is entirely feasible that children born within hospitals and other forms of medical facility are ceteris paribus more at risk of death than those born at home or in another non-medical environment as they represent a more risky birth in the first place.

In just over two-thirds (68.47%) of cases, women have a medical clinic in their local community. If we consider specific pregnancy related clinic visits, we observe that in households where no child mortality occurred over 93 percent of women regularly attended a clinic during their pregnancy. This rate is similar for households where child mortality did occur, with 90 percent of women regularly attending a clinic. Moreover, even where women do not have a clinic in their community, this does not seem to affect attendance at a clinic during pregnancy. Only 58 women who do not have a clinic in their area failed to attend a clinic regularly during pregnancy. It is more surprising that there were 197 women who have a clinic in their local community, but who chose not to regularly attend it throughout their pregnancy. I cannot exclude the possibility that these women did not attend the clinic because they were not aware of its existence.

[126] This is one such example within the proximate determinants framework.

Table 4.3.7 summarises the place of delivery for women. Hospitals are the most common place of delivery for women in Malawi, accounting for 44 percent of all births For births which take place in a hospital, 72 percent of children are delivered by a nurse, 14 percent by a doctor and 12 percent by a midwife. For women who have not experienced child mortality around 44 percent of women deliver in a hospital and a slightly smaller percentage (38 percent) of women deliver in a hospital where there has been at least one child mortality.

Table 4.3.7 Place of delivery

Place of delivery	Households experiencing no childhood mortality		Households experiencing a childhood mortality	
Hospital	*Frequency*	*Percent*	*Frequency*	*Percent*
Yes	*1444*	*44.49*	*200*	*38.10*
No	*1802*	*55.51*	*325*	*61.90*
Total	3246	100.00	525	100.00
Clinic				
Yes	*495*	*15.25*	*74*	*14.10*
No	*2751*	*84.75*	*451*	*85.90*
Total	3246	100.00	525	100.00
At home				
Yes	*943*	*29.05*	*192*	*36.57*
No	*2303*	*70.95*	*333*	*63.43*
Total	3246	100.00	525	100.00
Traditional birth attendant				
Yes	*323*	*9.95*	*52*	*9.90*
No	*2923*	*90.05*	*473*	*90.10*
Total	3246	100.00	525	100.00

Source: IHS2 *A small number of women indicated that they gave birth on the way to the hospital or "other". These are excluded from the Table due to the small number of cases concerned.

The second most common place for delivery is at home and accounts for 30 percent of births occurring in the last 2 years. Around 15 percent take place in a clinic, 10 percent with traditional birthing attendants and a very small number of births take place on the way to a hospital and in other facilities. Where a child is born at home, the delivery is normally not assisted by any trained medical personnel and instead the baby is delivered by the mother or another family member in over 90 percent of cases. Therefore, it is not unexpected that the rate of mortality is also relatively high in cases where the child was delivered at home.

Within the IHS2 there are some additional questions about the households' access and their perception of health services which may be relevant in terms of affecting the likelihood of child mortality. A new module was introduced in 2004 to examine households' subjective assessment of their well-being. There are two particular topics of interest which *may* be relevant in estimating the determinants of child mortality in Malawi; food consumption over the past month and standards of health care the household members receive. Such factors act as a proxy for investment in health which as Grossman (1972) shows increases the stock of health for an individual. Consequently, I would expect that improvements in food consumption and better health care would reduce the probability of experiencing mortality.

4.3.4 Perception of health services

Adequate nutrition is important for a mother to produce a healthy baby in the first instance but also to be able to provide nutrition via breastfeeding. Household heads were asked to evaluate whether or not the household members have had adequate food consumption over the past month. This is not an ideal indicator of nutrition as it provides no information about the nutrition of individuals (and my primary interest is the mother and any children). Ideally, I would like to know whether the mother has had adequate food (although more specifically nutrition) over the period just prior to conception, during pregnancy and for the following six months (assuming breastfeeding) along with details about the child's nutrition. However, this information is not available and so this is the next best alternative.

Only 36 percent of households[127] indicated that their food consumption over the past month had been adequate and 57 percent indicated that their food consumption had been less than adequate. An examination of the two types of household[128] does not indicate any large difference between them, in terms of their perception of food consumption. I would expect food consumption to be negatively related to the number of deaths: if the household has adequate nutrition the risk of child mortality is likely to be reduced.

[127] In my selected sub sample.
[128] Those households who have experienced a child mortality and those that have not.

In addition, households also provided an indicator about their perception of the standard of health care they have received. Again, this refers to all household members rather than specifically the mother or the child/children but it should not automatically be disregarded as it is an alternative measure of health care services. Just because someone has a medical clinic close by it does not mean that they will use it, as it may be too expensive. Moreover, just because someone accesses a medical service this does not mean that it is of a good quality.

Just over one third of households where a mother had given birth in the last 2 years indicated that their overall standard of health care received has been adequate whilst 60 percent indicated that the standard of health care they received was less than adequate. Again, the perception of households' standard of health care does not appear to differ substantially between the two types of household. I would expect that households which experience good standards of health care would be less likely to experience child mortalities as they seek to raise the stock of health for each child.

4.3.5 Under-five health interventions

Another additional module added to the IHS2 was a child anthropometry module, for children aged 6 months to five years of age. Mothers were asked whether their child had participated in a nutrition programme and an under-five clinic. Clearly households where all children have died in the past 2 years and/or where the household has only surviving children less than 6 months old will not be included in this module. However, it is possible to examine the behaviour of 84 percent of households where a mother had given birth in the last two years and had at least one living child aged 6 months to five years old. It is possible that being enrolled in an under-five clinic could be indicative of a households' general attitude towards child health. Households which undertake investments in a child's health would be expected to experience less child mortality as a child is being regularly monitored for signs of ill-health and treatment may be provided more readily than in cases where a child is not attending a clinic.

If I first consider the case where no child mortalities have occurred, I observe that just over a fifth of households who have at least one child aged under five did not enrol any of their children in an under-five clinic. A further fifth of households enrolled half of their children and 54 percent enrolled all of their children in a clinic. Turning to households who have experienced child mortality (but who have surviving children) I observe that almost a third did not enrol any of their children in a clinic which is somewhat higher than in non-affected households. There does not seem to be a particularly distinctive pattern between the two types of household or indeed even within households where child mortality has occurred and there are surviving children.

In section 4.2, I observed that a large number of deaths in Malawi can be attributed to malaria. Therefore, it is useful to know whether households make use of any preventative measures. Just over half of all households within my sample do not use mosquito nets at all for any household members. Where nets are used, there is a large difference in the use of mosquito nets between households who have and who have not experienced child mortality. Forty-one percent of households who have not experienced under-five mortality ensure that all children aged under five sleep under a mosquito net when mosquitos are present. In comparison, those households which have experienced at least one under-five mortality, ensure that in only 27 percent of cases all children aged under-five sleep under a net. I anticipate that the use of mosquito nets should be negatively related to the number of deaths experienced, as malaria is one of the biggest killers of children in Malawi.

4.3.6 Sanitation and water

Access to clean water is a key factor in reducing a child's risk of disease. Just under a fifth of households in the sample have access to piped water. In almost half of all households the source of drinking water is from a pump or protected spring reservoir and just over a third get their water from an unprotected well, spring or reservoir. Households that have experienced mortality do not have significantly different water sources to those who have not experienced mortality. Households with a clean/protected water supply should experience a lower risk of catching water borne disease and therefore the number of mortalities in these households should be lower. There is no additional information with the survey to indicate whether any steps are taken to try to ensure the safety of drinking water such as boiling water before drinking, or using purification tablets.

In terms of access to sanitation, approximately 82 percent of households have access to a latrine and this does not particularly differ between the two types of household. Only a tiny percentage of households indicate that they have access to a flushing toilet (less than 2%) in both cases. A small percentage of households have no access to sanitation facilities at all: 16% of households where there has been no mortality and 17% of households where there has been at least one mortality, respectively. I expect that mortality will be positively associated with households where sanitation is limited as children may be exposed to unhygienic conditions. Whilst only a very small percentage of households have flushing toilets, all other things being equal these households would be expected to experience the fewest number of mortalities as the risk of disease is lower.

In addition, disposal of general rubbish may also have an impact on a child's exposure risk to disease. Households reported that the most common form of waste disposal was via a collection from a rubbish pit in 57.9% of cases, followed by a communal heap 17.5% and in almost a fifth of cases (19.4%) there were no rubbish disposal facilities available to the household. Similar percentages of households use the same forms of rubbish disposal whether they have experienced child mortality or not and therefore there do not seem to be substantial differences between the two groups. Households which have no rubbish disposal facilities are likely to be most at risk as waste is not removed. Therefore, the possibility of child mortality for these households would be higher than for those with a collection from a rubbish pit where children are not in the vicinity of waste.

4.3.7 Household income and activity

The relationship between income and health is not a simple one, as outlined in *Chapter 2* and at the beginning of this chapter. Increased income gives people the opportunity to invest in their health; via better nutrition and access to medical treatment. Children of households where income is higher *may* be expected to have better health as the household is able to make health investments for the child. By including an indicator for household income, I can explore whether there is any evidence to suggest that increased household income might reduce the expected number and probability of experiencing under-five and infant mortality.

Table 4.3.8 Households' participation in different income earning activities

Type of activity	Households experiencing no childhood mortality		Households experiencing at least one child mortality	
	Frequency	*Percentage*	*Frequency*	*Percentage*
Dry season agriculture	*1915*	*36.81*	*230*	*43.81*
Rain-fed agriculture	*2904*	*89.46*	*490*	*93.33*
Rental income from land	*143*	*4.41*	*25*	*4.76*
Tobacco production	*492*	*15.16*	*95*	*81.90*
Livestock	*2129*	*65.59*	*352*	*67.05*
Self-enterprise	*113*	*34.29*	*181*	*34.48*
Tree crop production	*1240*	*38.20*	*219*	*41.71*
Other income sources	*224*	*6.90*	*23*	*4.38*

Source: IHS2

As around 30% of all households[129] do not participate in some form of waged employment, it is useful to examine other aspects of household activity when considering a measure of household income. Households may not generate a large income from waged employment but may have access to food or other resources directly from agricultural activities which would be expected to have a positive impact on household welfare and the stock of health for each child. A household which grows a range of foodstuffs may not appear to have a high income in principle but it may actually be able to more than adequately provide for its household members directly or via the sale of produce in the market.

[129] All households refers to each household which was surveyed in IHS2 and not my subsample.

Households within my sample reported several different sources of income in addition to wage income (either from formal employment or ganyu[130] activities); rental income from land, rain-fed and dry season agriculture,[131] livestock, tree crop and tobacco production, enterprise and a small selection of other sources such as pensions and savings.

As Table 4.3.8 shows, a large percentage of households have participated in income generating agricultural activities during the last cropping season, with a slightly higher percentage of households who have experienced child mortality engaging in agricultural activities[132] compared to those households which have not experienced any child mortality. However, whilst households may be able to generate revenue from their activities, it is the net value of any income that is of real interest. For rain-fed and dry season agriculture, households provide information about the costs associated with each type of production and it is therefore possible to estimate the net income received from these sources. Tobacco earnings are reported net of costs and so again net income from this source can be observed as can whether the household made a profit or a loss from any enterprises.

Households with higher levels of income should be more able to provide an environment which is conducive to good health, by providing health investments such as better nutrition. Furthermore, high income households may display desirable characteristics for other independent variables of interest such as education which are also negatively associated with child mortality as outlined in *Chapter 2*. In high income households, members are more likely to have a higher level of education on average and this in turn enables them to get a better paid job. Moreover, a more educated mother (or even more generally more educated household members) may make more informed decisions about a child's health in terms of investments and thus these characteristics serve to reduce the risk of mortality.

[130] This is a form of casual labour.
[131] In the last cropping season.
[132] All types of agriculture related activities including dry season, rain-fed, tobacco production, livestock and tree crop production.

In addition to household income, I also include an indicator for labour force participation for each mother. Data from the IHS2 indicates whether or not someone worked in the 12 months prior to the survey in both formal and informal (ganyu) employment, what type of occupation they were employed in, the type of employer and for how long they were working in a particular job. As I observed in *Chapter 3,* only a relatively small proportion of all individuals participated in any form of waged employment in the last year and only 188 individuals when I consider only women who have given birth in the last 2 years. If I also include casual employment, 1,246 women have worked for at least one day in the last year. This represents approximately a third of all women who have become mothers in the last two years. The rate of labour force participation is in reality not unexpected; women who have recently had children or have young children are very likely to be at home caring for their children rather than being outside the home working. Given the very small number of individuals working in any form, I use a simple dummy indicator for female labour force participation which combines those working in formal and informal employment against the alternative of not working for a wage. This indicator does not differentiate how long women have worked within the past year in any way given the small numbers involved. I note that of those women who are consequently labelled as employed, a number have had both formal and informal employment in the last year.

4.3.8 Limitations of the variables

One potential issue of concern is that the number of deceased children could be underreported if still births are not considered to be child deaths. The survey does not specifically refer to deaths following a live birth and so some households may view such deaths differently to others; some may count them, whilst others don't. There is unfortunately no additional information within the survey that could be used to identify still births. One should also not forget that death and particularly the loss of a child is always a highly emotive subject. As such, it is likely in the absence of formal records which would verify the information provided, that the number of deaths may not be reported accurately by mothers. Hence the number of infant and child mortalities might be underreported.

Moreover, the recall period is relatively short at two years but some households may not accurately report the age of the deceased and so the number of deaths could be misreported which would lead to measurement error, as this could lead to the inclusion or exclusion of children. Provided the measurement error is random and not associated with the independent variables, it shouldn't cause bias in the regression estimates. However, there is a possibility that this type of measurement error could be associated with some of the independent variables, particularly education level. One would expect that more educated women would keep better records/have a better memory about particular events and as such they may report child deaths more accurately than those with less education. In addition, I observe that a significant percentage of women do not have any education at all and therefore there is a possibility that misreporting could lead to bias in my estimators. This is a concern but in the absence of any additional means of verification, I cannot address this issue.

During the selection of the sample, it became clear that there are a number of reporting errors within the survey. In particular the number of reported births from the health module did not coincide with the information reported within the household roster and on deceased children. Using information from the household roster and reported deaths combined with the information provided about births in the past two years, I have been able to verify 3,771 households as having had at least one child in the past 2 years.

A large number of households[133] do not report the level of education of the mother and as previously stated approximately thirty percent of all households do not have anyone in the household participating in any form of waged employment. Therefore, a large number of households would not be included when estimating my regression coefficients; a similar issue also arises when considering alternative measures of household income.

The coefficient on log income used within my regression estimates is determined by those households which reported an income; households which do not have an income would therefore not be included. By adding in a dummy variable for log income, I am not constrained to only those households which report an income so I can still consider

[133] 1,114 households within my sample, which represent 30% of the sample.

the whole sample. In *Chapter 3,* I outlined how the education level measure was created. In the regressions, I use either a dummy for whether or not the person has any education, or a set of dummies for each level of education, with no education as the omitted benchmark level.

Almost a third of women did not report any level of education at all, therefore the information is missing. Omitting these observations would result in a large decrease in sample size. One potential option for education would be to recode the missing values to zero to represent "no education" to maintain my sample. However, given that I cannot be sure whether the values do really represent "no education" I have added a dummy for "missing education" instead. Using this technique, the coefficients on the other variables apply to all observations, and the coefficients on the education variables apply only to those observations for which we observe education levels.[134]

4.3.9 Model choice

In Table 4.3.1, I observe that there are a small number of households which experience more than one childhood death. This enables me to consider a count model such as the Poisson or a more general model with less strict assumptions such as the negative binomial. This is in addition to the basic binary choice model such as a probit or logit.

Using a binary choice model, I seek to find the probability that the event (in this case a child or infant mortality) occurs conditioned on the set of my explanatory variables which will include variables such as characteristics about the mother, their place of residence, health services and household income.

In a linear model such as OLS, the coefficient on each independent variable can be easily interpreted to provide an estimate of the effect that a small change in that variable will have on the dependent variable. With a binary response model, the magnitudes of the coefficients on each of the dependent variables are not that informative, as all they tell me is the direction of the effect on the outcome and whether or not the independent

[134] This is known as "dummy variable adjustment".

variable is significant or not. Viewing the marginal effects[135] of each independent variable is of more interest as it shows me how a small change in the independent variable will affect the probability of the outcome occurring (Wooldridge, 1999).

Whether a household experiences an under-five or infant mortality can also be thought of as a count outcome. A count model is useful when there are usually only a few possible outcomes (a positive integer). A count variable is not continuous and therefore a normal distribution is not suitable. Instead, in a count model, the expected outcome can be zero or a positive integer and therefore the Poisson distribution is more appropriate.

One of the restrictions of the Poisson distribution is that its variance is equal to its mean. In practice this is often not the case and leads to two problems known as overdispersion and underdispersion. Overdispersion occurs when the variance is larger than the mean. Underdispersion occurs when the variance is less than the mean, although the former is the more common problem experienced in practice (Wooldridge, 1999). An alternative to use in such a case is the negative binomial model; the assumptions of this model are less restrictive and it may therefore be more appropriate. I tested for this in my estimations, and found no evidence to support the proposition that a Poisson model is inappropriate for the data used. Therefore Poisson and probit results are reported within the next section.

4.4 Regression results

As already noted, my dependent variable will be either a count variable or a binary indicator variable. In this section, I present a series of specifications which examine the determinants of child and infant mortality in Malawi. The first set of models presented will consider the characteristics of the mother, geographical location and some basic health services indicators in a broad sense. The variables used throughout the models presented in this section are summarised in Table 4.3.9.

[135] I am using the mfx command within Stata 10.

Table 4.3.9 Summary list of variables

Variable	Variable definition
logy	log(earned income) from wages, ganyu labour and in-kind sources
logydummy	dummy: equal to 1 if household income is missing
births	number of births within the past 2 years
rural	dummy: equal to 1 if household is based in the rural area
north	dummy: equal to 1 if the household is based in the northern region
centre	dummy: equal to 1 if the household is based in the central region
educated	dummy: equal to 1 if the mother possesses at least primary level education
educdummy	dummy: equal to 1 if education is missing
employed	dummy: equal to 1 if the mother was employed in either formal or informal activities in the last 12 months
primary	dummy: equal to 1 if the highest level of education achieved by the mother is primary level
secondary	dummy: equal to 1 if the highest level of education achieved by the mother is secondary level
tertiary	dummy: equal to 1 if the highest level of education achieved by the mother is tertiary level
clinic	dummy: equal to 1 if the mother delivered her baby at a clinic
home	dummy: equal to 1 if the mother delivered her baby at home
attend	dummy: equal to 1 if the mother delivered her baby at a birth attendant centre
on way	dummy: equal to 1 if the mother delivered her baby on the way to the hospital
other	dummy: equal to 1 if the mother delivered her baby in some other way
doctor	dummy: equal to 1 if the mother was assisted by a doctor during birth
midwife	dummy: equal to 1 if the mother was assisted by a midwife during birth
TBA	dummy: equal to 1 if the mother was assisted by a traditional birthing assistant during birth
self	dummy: equal to 1 if the mother was assisted by only herself or another family member during birth
go clinic	dummy: equal to 1 if the mother regularly attended a clinic during her pregnancy
allnet	dummy: equal to one if all of the children under-five in the household sleep under bednets
somenet	dummy: equal to one if some of the children under-five in the household sleep under bednets
pipewater	dummy: equal to 1 if the household gets their water from a piped source
wellspringres	dummy: equal to 1 if the household gets their water from a well, spring or reservoir
flush	dummy: equal to 1 if the household has a flushing toilet
notoilet	dummy: equal to 1 if the household has no access to a toilet
bin	dummy: equal to 1 if the household has a bin to dispose of their rubbish
heap	dummy: equal to 1 if the household has a heap to dispose of their rubbish
no rubbish	dummy: equal to 1 if the household has no access to rubbish disposal

Based on the discussion in *Chapter 2* I hypothesise that an increased use of health services along with better living conditions will be negatively associated with child and infant mortality. In addition, I also will examine the proposition that increases in household income will reduce the expected probability of child and infant mortality and the expected number of child and infant mortalities. The final sets of independent variables used in the regression models are associated with characteristics about the mother and her location. I hypothesise that an increase in education will be negatively associated with child and infant mortality as will living in the urban area based on the discussion in *Chapter 2.*

In the tables that follow, the raw coefficients are presented in the first column and in the second column the marginal effects are reported. The sign on the coefficient indicates the direction of the effect but for interpretation I examine the marginal effects. In each of the tables the marginal effects have been calculated at the mean for the continuous variables. For dummy variables the interpretation is slightly different; the marginal effect is the change in moving from one state to another where D=0 and then when D=1.

Table 4.4.1 presents the results from some initial regressions using a Poisson and a probit for child mortality. The independent variables included a measure of income,[136] some characteristics about the mother, location and the number of births that the mother has had during the past 2 years and some indicators of health services. Given that under or overdispersion can be a concern when using a Poisson model, I first examine the mean and the variance of the count variable child mortality used in the Poisson estimation. The mean value is 0.154 and the standard deviation reported is 0.401 which gives a variance of 0.161. Therefore, the mean and variance of this indicator are reasonably close. A further check is to examine the results of the negative binomial for the same model and compare the coefficients between the Poisson and the negative binomial. I do not report the estimation results from the negative binomial; however, I can confirm for the reader that the results from the two estimations are very similar.

4.4.1 Income and mortality

As outlined very briefly in the previous section, the sign of each of the marginal effect coefficients is indicative of the direction of the effect that the independent variable has on the expected number (or expected probability in the case of a probit) of under-five/infant mortalities. As a large number of households do not report any form of waged income a dummy is included to ensure that all households are included and then

[136] Specifically, income from employment in the last year.

treats all of the coefficients on the other independent variables as equal for households whether or not they receive any form of waged income.

For robustness purposes the models presented within this section are also run using different measures of household income such as the income received from rain-fed and dry season agriculture. I therefore explore whether the results are sensitive to the measure of income used. The results from such alternative specifications[137] also indicate that the expected number and probability of experiencing an under-five or infant mortality is not influenced by household income.

Specifically, from looking at the marginal effects of income in more detail from Table 4.4.1 I observe that the coefficient on income in the Poisson model is negative, very small and insignificantly different from zero. Even if (log) household income were to increase by 50 percent (at the mean), which would be a relatively large change, there would be less than half a percent decrease in the number of child mortalities occurring. In the probit model, the probability of a child mortality occurring is also barely affected by such a large change in income (in the case of the probit model, the coefficient is actually positive but also insignificantly different from zero). The standard errors for these marginal effects are also rather small (less than 0.005 in both cases), implying that the marginal effects are fairly precisely estimated. The estimations suggest that, holding other characteristics constant (including, importantly, education – see below), income does not have an independent impact on the expected number and probability of child mortality occurring within Malawian households.

Following the results presented in Table 4.4.1 I also examined the Poisson and probit model results using infant mortality as the dependent variable in Table 4.4.2. The results are very similar in that they indicate that increased household income will reduce infant mortality but again this is not statistically different from zero. There is no evidence to support the proposition that increased household income reduces under-five or infant mortality within Malawi given the results presented and using alternative measures of household income.

[137] Not presented.

In the majority of the specifications there are however a number of variables that appear to be consistently relevant in terms of affecting the number and/probability of child and infant mortalities; the number of births that occurred within the past two years, whether the household resides in the rural area or a specific region and whether the mother has any education.

4.4.2 The importance of education

Education may be examined in a regression model in a number of ways: a simple dummy of whether or not the mother has any education, a series of dummies for different levels of education, years of education, or an indicator of literacy. The results presented in Table 4.4.1 and Table 4.4.2 consider education as a simple dummy (whether or not the mother has any formal education) and in Table 4.4.3 education level dummies are used. I return to the use of a simple education dummy for the remainder of the chapter given that a relatively small number of women in the data possess any education at all.

I also considered the number of years of education that each mother reported. This measure is based on when they first attended school and when they then left school to calculate the number of years they have spent in school. However, this approach is not ideal as an individual could spend many years in education which may not actually translate into being more academically able. Moreover, using the number of schooling years in a regression model[138] such as those presented in Table 4.4.1 and 4.4.3 indicates that the number of schooling years does not have a statistically significant impact on under-five mortality or infant mortality which would support the notion that the length of time one spends in education does not necessarily translate to being educated.

Another alternative that could be considered is an indicator of literacy. Respondents were asked about their ability to read a one page letter in Chichewa, English and in any other language.

[138] Not presented.

Table 4.4.1 Results from Poisson and probit with under-five deaths as the dependent variable and a simple education dummy

	Poisson		Probit	
	Coefficient	Marginal effect	Coefficient	Marginal effect
logy	-0.033 (0.027)	-0.004 (0.003)	0.003 (0.020)	0.001 (0.004)
logydummy	-0.370 (0.247)	-0.037 (0.023)	-0.079 (0.186)	-0.014 (0.033)
north	-0.357 (0.137)**	-0.034 (0.011)***	-0.281 (0.094)***	-0.045 (0.013)***
centre	-0.023 (0.085)	-0.002 (0.009)	0.010 (0.064)	0.002 (0.012)
educated	-0.460 (0.174)**	-0.042 (0.013)***	-0.282 (0.106)***	-0.045 (0.015)***
educdummy	-0.082 (0.085)	-0.009 (0.009)	-0.089 (0.065)	-0.016 (0.011)
employed	-0.1223 (0.091)	-0.013 (0.009)	-0.1334 (0.069)*	-0.024 (0.012)**
goclinic	-0.304 (0.143)**	-0.037 (0.020)*	-0.233 (0.112)**	-0.048 (0.026)*
clinic	0.001 (0.123)	0.0001 (0.013)	0.054 (0.086)	0.010 (0.016)
home	0.160 (0.207)	0.017 (0.023)	0.297 (0.192)	0.058 (0.041)
other	0.094 (0.623)	0.010 (0.072)	0.200 (0.431)	0.041 (0.099)
attend	0.429 (0.250)**	0.054 (0.038)	0.482 (0.217)**	0.110 (0.060)
on way	0.110 (0.563)	0.012 (0.066)	-0.139 (0.407)	-0.023 (0.062)
doctor	0.230 (0.151)**	0.036 (0.021)*	0.171 (0.110)	0.034 (0.024)
midwife	-0.010 (0.159)	-0.001 (0.017)	0.051 (0.113)	0.010 (0.022)
TBA	-0.630 (0.219)***	-0.056 (0.017)***	-0.559 (0.200)***	-0.083 (0.024)***
self	0.025 (0.208)	0.003 (0.022)	0.064 (0.195)	-0.011 (0.034)
births	1.762 (0.058)***	0.186 (0.008)***	1.769 (0.084)***	0.323 (0.018)**
allnet	-0.548 (0.089)***	-0.055 (0.009)***	-0.400 (0.065)***	-0.069 (0.011)***
somenet	-0.780 (0.349)***	-0.059 (0.018)***	-0.619 (0.208)***	-0.077 (0.016)***
pipewater	0.045 (0.115)	0.005 (0.013)	0.041 (0.086)	0.008 (0.016)
wellspringres	0.011 (0.083)	0.001 (0.009)***	-0.016 (0.064)	-0.003 (0.012)
flush	-0.681 (0.356)*	-0.053 (0.020)	-0.584 (0.250)	-0.073 (0.020)***
notoilet	-0.085 (0.104)	-0.009 (0.010)	-0.058 (0.080)	-0.104 (0.014)
bin	-0.105 (0.368)	-0.011 (0.035)	-0.065 (0.241)	-0.011 (0.041)
heap	-0.029 (0.105)	-0.003 (0.011)	-0.072 (0.080)	-0.013 (0.014)
norubbish	0.056 (0.099)	0.006 (0.011)	0.040 (0.075)	0.007 (0.014)
Wald chi2 (27)	1178.42		527.65	

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses. N = 3771, Pseudo R-squared for Poisson = 0.1841 and 0.2239 for the probit.

However, within the question there is nothing to qualify how well someone can read whereas a formal qualification is at least indicative of reaching a nationally recognised standard. Therefore, my preference is to use dummy variables to represent levels of education achieved.

Given that a large number of women do not report their any level of education, I have included a dummy variable for non-reporting[139] to ensure that the sample of 3,771 households is maintained. The alternative is to replace such missing values of education[140], to drop such cases, or insert the sample mean level of education. In the simple case shown in Table 4.4.1 women that reported that they had no education are coded as 0 and women who reported that they achieved a formal qualification which could be a primary school leavers certificate or above are coded as a 1, for the variable labelled “educated”. In Table 4.4.2, I then consider the different levels of education achieved; primary, secondary and tertiary and create a dummy variable for the categories, excluding no education as the base category and again add a dummy for non-reporting to ensure that all cases are used when the level of education is missing

Table 4.4.1 shows that possessing at least some education has a positive and statistically significant impact in reducing under-five mortality, reducing the expected number of child mortalities by 4.2 percent and the probability of experiencing a child mortality by just over 4 percentage points.

Given that the under-five mortality rate is 14 percent within my sample, a reduction of four percentage points is relatively large, and this is assuming that the mother only has at least some level of formal education. In Table 4.4.2 I consider infants and observe that whilst the marginal effect of possessing some education is still negative it is a smaller effect than that observed for under-fives. A possible reason for this may be associated with the cause of death which I explore in Table 4.4.5. Notably, in the case of under-fives, possessing some education reduces the probability of experiencing

[139] This dummy is labelled as “educdummy”.
[140] One may assume that this could be “no education”. However, when respondents were asked about their level of education this was an option they could select. Therefore I have opted not to replace such values and use dummy variable adjustment instead and also run a separate set of regressions which use the smaller sample for comparison.

mortality by 3.5 percentage points in the case of illness related deaths. This falls to one percentage point in the case of non-illness related deaths. For infants the effect of education is even smaller in both cases. A well-educated mother may make more informed decisions about the care of her child such as ensuring adequate nutrition, immunisation and seeking medical attention when necessary. Such behaviour could reasonably be expected to reduce exposure to the risk of illness and in turn reduce the probability of illness related death. In contrast, Table 4.3.4 shows that a significant proportion of non-illness related deaths are attributed to birth complications, something that a mother is unlikely to have much control over.

If I examine the level of education in a little bit more detail, I observe in Table 4.4.3 that in a household where a mother has primary education compared to a household where the mother has no education, the effect is to reduce the expected number of under-five mortalities by 5.2 percent and this result is statistically significant at the 5% level. In the right hand side of Table 4.4.3 the results for the probit model show that having primary education reduces the probability of experiencing child mortality by more than 5 percentage points compared to a household where the mother has no education. A mother possessing secondary level education is predicted to experience less under-five mortality, but by a smaller margin (just over 2 percent) and this result is not statistically significant.

The results from Table 4.4.3 demonstrate that increased education reduces the number of under-five mortalities and a similar result is also found for infants. However, it would appear that the greatest gains can be made by ensuring that mothers have achieved primary level education. I note that the marginal effect of possessing tertiary level of education is very large, reducing the expected number of under-five mortalities by over 10 percent. However, one must also appreciate that in my sample of 3,771 women, only four women possess tertiary level education. Therefore, despite there being statistically significant evidence to support that tertiary education can reduce the expected number of under-five mortalities by a relatively large degree; in practice so few women possess tertiary level education that its practical significance cannot be reliably determined. Moreover, the probit results drop tertiary level education as no women with tertiary level education experienced any mortality.

In sum, the real opportunity appears to lie in assisting women to complete primary education (which is an important issue with many other areas of development literature), which is both statistically and practically significant in terms of reducing the expected number and probability of experiencing a under-five mortality and to a lesser degree infant mortality.

4.4.3 Location

Households report the region of Malawi that they are living in: north, centre and south. The majority of births and deaths occur in the south and therefore this is my omitted dummy for comparison. The results presented in Tables 4.4.1-4.4.5 show that living in the north of the country is negatively associated with under-five and child mortality.

Table 4.4.1 shows that living in the north reduces the number of under-five mortalities by 3.4 percent and the probability of experiencing child mortality by 4.5 percentage points compared to households based in the south. Again, based on the overall under-five mortality rate of 14 percent, being located in the north relative to the south appears to play a substantial role in predicting the survival outcomes for children living in these areas. Living in the centre does not appear to significantly affect the number or probability of under-five mortalities in any specification. It could be that there are unobserved factors which are associated with each of the regions such as the public health programme within each area which may explain the large differences experienced across the regions. It could be, for example, that the northern region has more investment in public health care provisions such as doctors or more access to medication relative to the south. However, without additional information to control for potential differences it is not possible to explore this possibility in any additional detail.

I also examine the type of location: rural or urban. Taking the approach presented in Tables 4.4.1 and 4.4.2 but adding in an indicator for with whether the residence is rural or urban, I observe in Table 4.4.4 that living in the rural area is positively associated with under-five mortality. However, this finding is not statistically significant.

Table 4.4.2 Results from Poisson and probit with infant deaths as the dependent variable and a simple education dummy.

	Poisson		Probit	
	Coefficient	Marginal effect	Coefficient	Marginal effect
logy	-0.056 (0.035)	-0.004 (0.002)	-0.012 (0.023)	-0.002 (0.003)
logydummy	-0.516 (0.320)	-0.030 (0.017)*	-0.153 (0.216)	-0.018 (0.024)
north	-0.507 (0.163)***	-0.027 (0.007)***	-0.308 (0.103)***	-0.032 (0.009)***
centre	-0.340 (0.114)***	-0.021 (0.007)***	-0.193 (0.073)***	-0.023 (0.008)***
educated	-0.395 (0.208)*	-0.022 (0.010)**	-0.201 (0.114)*	-0.022 (0.011)**
educdummy	-0.268 (0.116)**	-0.016 (0.007)**	-0.178 (0.075)**	-0.021 (0.008)**
employed	-0.081 (0.120)	-0.005 (0.007)	-0.079 (0.077)	-0.010 (0.009)
goclinic	-0.468 (0.184)**	-0.037 (0.018)**	-0.332 (0.123)***	-0.051 (0.023)**
clinic	-0.210 (0.161)	-0.012 (0.009)	-0.080 (0.098)	-0.009 (0.011)
home	-0.155 (0.242)	-0.010 (0.015)	-0.019 (0.174)	-0.002 (0.021)
other	0.162 (0.606)	0.011 (0.045)	0.117 (0.407)	0.016 (0.059)
attend	0.016 (0.307)	0.001 (0.020)	0.045 (0.209)	0.006 (0.027)
on way	-0.054 (0.554)	-0.003 (0.033)	-0.016 (0.185)	-0.002 (0.047)
doctor	0.429 (0.196)**	0.033 (0.018)*	0.217 (0.123)*	0.031 (0.020)
midwife	-0.009 (0.207)	-0.001 (0.013)	0.040 (0.131)	0.005 (0.017)
TBA	-0.171 (0.266)	-0.010 (0.015)	-0.116 (0.185)	-0.013 (0.020)
self	0.318 (0.246)	0.022 (0.019)	0.182 (0.178)	0.024 (0.025)
births	1.835 (0.070)***	0.117 (0.007)***	1.429 (0.079)***	0.176 (0.012)***
allnet	-0.778 (0.121)***	-0.047 (0.007)***	-0.541 (0.074)***	-0.061 (0.008)***
somenet	-1.051 (0.463)**	-0.043 (0.011)***	-0.633 (0.249)**	-0.049 (0.011)***
pipewater	0.165 (0.142)	0.011 (0.010)	0.126 (0.092)	0.016 (0.013)
wellspringres	-0.075 (0.109)	-0.005 (0.007)	-0.061 (0.072)	-0.007 (0.009)
flush	-0.495 (0.387)	-0.025 (0.015)	-0.408 (0.270)	-0.037 (0.017)**
notoilet	-0.137 (0.143)	-0.008 (0.008)	-0.121 (0.092)	-0.014 (0.010)
bin	0.203 (0.412)	0.014 (0.032)	0.119 (0.245)	0.016 (0.036)
heap	0.007 (0.132)	0.0004 (0.008)	-0.068 (0.088)	-0.008 (0.010)
norubbish	0.162 (0.126)	0.011 (0.009)	0.103 (0.082)	0.013 (0.011)
Wald chi2 (27)	933.01		418.59	

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses. N = 3771, Pseudo R-squared for Poisson = 0.1859 and 0.1973 for the probit.

Table 4.4.3 Results from Poisson and probit with under-fives deaths as the dependent variable and differing levels of education dummies

	Poisson		Probit	
	Coefficient	Marginal effect	Coefficient	Marginal effect
logy	-0.036 (0.027)	-0.004 (0.003)	0.002 (0.020)	0.0004 (0.004)
logydummy	-0.038 (0.247)	-0.037 (0.022)*	-0.086 (0.187)	-0.015 (0.033)
north	-0.343 (0.135)**	-0.032 (0.011)***	-0.274 (0.094)***	-0.044 (0.013)***
centre	-0.022 (0.085)	-0.002 (0.009)	0.009 (0.064)	0.002 (0.012)
primary	-0.636 (0.238)***	-0.052 (0.014)***	-0.971 (0.147)**	-0.055 (0.017)***
secondary	-0.226 (0.230)	-0.021 (0.020)	-0.176 (0.137)	-0.029 (0.020)
tertiary	-10.512 (0.278)***	-0.106 (0.005)***	dropped	
educdummy	-0.083 (0.085)	-0.009 (0.009)	-0.089 (0.065)	-0.016 (0.011)
employed	-0.123 (0.092)	-0.013 (0.009)	-0.135 (0.069)*	-0.024 (0.012)**
goclinic	-0.304 (0.143)**	-0.036 (0.020)*	-0.231 (0.112)	-0.048 (0.026)*
clinic	0.003 (0.123)	0.0003 (0.013)	0.056 (0.086)	0.011 (0.016)
at home	0.169 (0.207)	0.018 (0.023)	0.303 (0.193)	0.060 (0.041)
other	(0.114 (0.623)	0.013 (0.073)	0.208 (0.432)	0.043 (0.100)
attend	0.440 (0.250)*	0.055 (0.038)	0.488 (0.218)**	0.112 (0.060)*
on way	0.116 (0.561)	0.013 (0.066)	-0.137 (0.406)	-0.023 (0.062)
doctor	0.298 (0.151)*	0.035 (0.020)*	0.172 (0.110)	0.034 (0.024)***
midwife	-0.013 (0.159)	-0.001 (0.016)	0.050 (0.113)	0.009 (0.022)
TBA	-0.638 (0.219)***	-0.056 (0.016)***	-0.563 (0.200)***	-0.084 (0.018)***
self/other family	0.021 (0.207)	0.002 (0.022)	-0.068 (0.195)	-0.012 (0.034)
births	1.767 (0.058)***	0.185 (0.008)***	1.772 (0.084)***	0.324 (0.018)***
allnet	-0.559 (0.090)***	-0.056 (0.009)***	-0.405 (0.065)***	-0.070 (0.011)***
somenet	-0.778 (0.348)**	-0.058 (0.018)***	-0.617 (0.208)***	-0.077 (0.016)***
pipewater	0.037 (0.116)	0.004 (0.012)	0.035 (0.086)	0.006 (0.016)
wellspringres	0.010 (0.083)	0.001 (0.009)	-0.016 (0.064)	-0.003 (0.012)
flush	-0.691 (0.352)**	-0.053 (0.019)***	-0.596 (0.251)**	-0.074 (0.019)***
notoilet	-0.084 (0.104)	-0.009 (0.010)	-0.057 (0.080)	-0.010 (0.014)
bin	-0.125 (0.367)	-0.012 (0.034)	-0.078 (0.240)	-0.014 (0.040)
heap	-0.028 (0.105)	-0.003 (0.011)	-0.072 (0.080)	-0.013 (0.014)
norubbish	0.057 (0.099)	0.006 (0.011)	0.040 (0.075)	0.007 (0.014)
Wald chi2 (29)	1739.09		530.64	

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses. N = 3771 for Poisson and 3767 for probit , Pseudo R-squared for Poisson = 0.1846 and 0.2240 for the probit.

Table 4.4.4 Results from Poisson and probit with under-five deaths as the dependent variable and the inclusion of a rural residence dummy

	Poisson		Probit	
	Coefficient	Marginal effect	Coefficient	Marginal effect
logy	-0.030 (0.027)	-0.003 (0.003)	0.005 (0.020)	0.001 (0.004)
logydummy	-0.344 (0.247)	-0.034 (0.023)	-0.061 (0.187)	-0.011 (0.034)
north	-0.353 (0.137)***	-0.033 (0.011)***	-0.279 (0.094)***	-0.045 (0.013)***
centre	-0.023 (0.086)	-0.002 (0.009)	0.010 (0.064)	0.002 (0.012)
rural	0.145 (0.140)	0.015 (0.013)	0.077 (0.099)	0.014 (0.017)
educated	-0.446 (0.175)**	-0.040 (0.013)***	-0.273 (0.106)***	-0.044 (0.015)***
educdummy	-0.082 (0.084)	-0.009 (0.009)	-0.090 (0.065)	-0.016 (0.011)
employed	-0.129 (0.091)	-0.013 (0.009)	-0.136 (0.069)**	-0.024 (0.012)*
goclinic	-0.304 (0.143)**	-0.037 (0.020)*	-0.233 (0.112)**	-0.048 (0.026)*
clinic	-0.007 (0.124)	-0.001 (0.013)	0.053 (0.086)	0.010 (0.016)
at home	0.145 (0.207)	0.016 (0.023)	0.291 (0.192)	0.057 (0.040)
other	0.079 (0.623)	0.009 (0.071)	0.194 (0.432)	0.040 (0.098)
attend	0.420 (0.249)*	0.053 (0.037)	0.479 (0.217)**	0.109 (0.060)*
on way	0.120 (0.552)	0.014 (0.066)	-0.138 (0.401)	-0.023 (0.061)
doctor	0.312 (0.152)**	0.038 (0.021)*	0.179 (0.111)	0.034 (0.024)
midwife	-0.012 (0.159)	-0.001 (0.017)	0.049 (0.114)	0.009 (0.024)
TBA	-0.631 (0.219)***	-0.056 (0.017)***	-0.560 (0.200)***	-0.083 (0.024)***
self	0.027 (0.207)	0.003 (0.022)	-0.064 (0.195)	-0.011 (0.034)
births	1.763 (0.058)***	0.186 (0.008)***	1.769 (0.084)***	0.323 (0.018)***
allnet	-0.537 (0.089)	-0.054 (0.009)***	-0.397 (0.065)***	-0.069 (0.011)***
somenet	-0.785 (0.350)**	-0.059 (0.018)***	0.620 (0.209)***	-0.077 (0.016)***
pipewater	0.084 (0.119)	0.009 (0.013)	0.064 (0.091)	0.012 (0.017)
wellspringres	0.010) (0.083)	0.001 (0.009)	-0.015 (0.064)	-0.003 (0.012)
flush	-0.696 (0.361)*	-0.054 (0.020)***	-0.583 (0.248)**	-0.073 (0.019)***
notoilet	-0.087 (0.104)	-0.009 (0.010)	-0.061 (0.080)	-0.011 (0.014)
bin	-0.053 (0.370)	-0.006 (0.037)	-0.037 (0.244)	-0.007 (0.043)
heap	-0.025 (0.105)	-0.003 (0.011)	-0.069 (0.080)	-0.012 (0.014)
norubbish	0.050 (0.099)	0.005 (0.011)	0.038 (0.075)	0.007 (0.014)
constant	-3.283		-2.734	
Wald chi2 (28)	1186.70		527.93	

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses. . N = 3771 Pseudo R-squared for Poisson = 0.1844 and 0.2241 for the probit.

When I consider infants[141] the Poisson results show that rural households are expected to experience 2 percent more deaths and this result is statistically significant at the 5% significance level. The probit results also confirm that those living in the rural area are more likely to experience an infant death by 2.2 percentage points. If the cause of death is also considered as in Table 4.4.5, it is shown that living in the rural area is positively associated with the probability of experiencing both an under-five and infant mortality in the case of non-illness related deaths, increasing the expected probability of mortality by 2.5 percentage points in both cases and significant at the 1% significance level.

Given that controls are in place for other factors such as maternity/antenatal care and sanitation the results in Table 4.4.5 suggest that all other things being equal, children living in the rural area are more likely to experience mortality that their equivalent urban counterparts.

However, the marginal effect of rural residence is statistically significant only in the case of non-illness related deaths. Once again, it is worth bearing in mind that the majority of non-illness related deaths are associated with birth complications. Whilst there are medical facilities such as clinics within the rural areas it could be that these facilities are not as well equipped as those in the urban area which the coefficient on rural residence is perhaps capturing.

As in the earlier discussion surrounding geographical region, it is not possible to observe differences within the healthcare provisions in more detail and therefore this possibility cannot be explored any further. Whilst the probability of experiencing an under-five and mortality is higher in the rural area, one must also remember that Malawi is predominately rural, with only 13 percent of households being located in the urban area as indicated in section 3.3.1. Location appears to demonstrate differences in the probability of mortality. Given that there are controls for the presence of health facilities it would suggest that there are other factors which need to be considered. One such factor if there were available information would be some measure of the quality of facilities across different areas.

[141] Results not presented.

4.4.4 Health and pregnancy related services

As previously outlined increased use of and ease of access to health services is likely to reduce under-five mortality. In addition I proposed that households' perception of health services and food consumption may highlight the importance of quality of health services and adequate nutrition. Their inclusion did not indicate that food consumption or quality of health services were statistically significant in any specification in terms of affecting child and infant mortality. Therefore the variables related to perception of health services, food consumption and delivery location have not been included in the results presented. Instead, the tables include the following health services variables: whether or not the mother attended a clinic regularly during her pregnancy, where she delivered her baby, and who delivered her baby.

Somewhat surprisingly, the location of where the mother delivered her baby was generally not significant in any specification. The exception to this is in cases where a baby was delivered at a traditional birthing centre, as shown in Table 4.4.3 and 4.4.4. In both cases the calculated marginal effect is very large but imprecisely estimated, with large standard errors and only just statistically significant at the 10 percent significance level. The calculated marginal effect in Table 4.4.4 suggests that children born in such a facility are almost 11 percentage points more likely to die than those born in a hospital. Given the overall rate of mortality is 14 percent, this result seems questionable. Moreover, the 95% confidence interval is very large indeed, ranging from -0.008 to 0.226. Relatively few women delivered their most recent child in such a centre and of these women a small percentage experienced mortality as shown by Table 4.3.7. Therefore, practically this result cannot really be viewed with any degree of confidence. The expected number and probability of experiencing under-five mortality is greater when the child is not delivered at a hospital, but this effect is not statistically different from zero. One may argue that children born in a hospital may have access to certain medical facilities and personnel that are not available in other facilities; particularly the home. However, a counter argument to this is that children born in hospitals represent more risky births in the first place and therefore the risk of mortality is greater for such births.

There is therefore a possibility that there is a selection problem in terms of choosing where delivery takes place. Women who have not had any children previously and who have not had complications throughout their pregnancy may not be concerned with seeking medical assistance for the birth of their child. It is possible that women only subsequently use a medical facility and/or a trained member of staff to deliver their child following complications at a previous birth, or if they have experienced difficulties during their current pregnancy. In such a case a woman who seeks medical assistance for these reasons may be at a higher risk of child/infant mortality from the outset.

Hence births taking place outside of hospitals may take place at home and other facilities as they are lower risk births and do not warrant hospitalisation.[142] The results presented within Table 4.4.1 indicate that births taking place outside of hospitals result in a higher incidence and probability of mortality. However, Table 4.4.2 suggests that infant mortality will be lower when the child is born at a clinic, at home or on the way to the hospital relative to those infants born within hospitals which would support the proposition that hospital births are perhaps linked to more risky births from the outset. Given that the indicators on where a child are born are only relevant to the most recently born child, I would have more confidence about the direction of the marginal effects presented in Table 4.4.2. However, the results in both Table 4.4.1 and 4.4.2 indicate that the findings related to the place of delivery are not statistically different from zero.

A similar pattern is observed when considering who delivered the most recently born child. The results from Table 4.4.1 suggest that in cases where a doctor delivered the child, the expected number of under-five mortalities is slightly higher than in cases where the child was delivered by a nurse and this result is statistically significant at the 10% level. A similar kind of argument to that presented above regarding location of delivery can be considered here with regard to who delivered. It could be the case that doctors are only used by women in cases where the complications are extreme and require more specialist assistance than by the usual route of a nurse. Thus, by this very nature the risk of mortality is higher. If this were the case then this could explain why

[142] A small number of women gave birth on the way to the hospital which would suggest that they would have given birth in hospital had they arrived before delivering their baby.

the sign on doctor is positive. This result is also found by Manda (1999) and he too proposes a similar idea: complicated births take place in hospitals and therefore the risk of mortality may be greater.

Table 4.4.2 shows that the probability of experiencing infant mortality is 3.1 percentage points higher in cases where delivery is undertaken by a doctor relative to those delivered by a nurse, but the result is not statistically significant. Another result of interest is that of the effect of a child being delivered by a traditional birthing assistant (TBA) in the case of under-five mortalities. In both the Poisson and probit results presented in Table 4.4.1, mortality is lower amongst those delivered by a TBA relative to a nurse and by a large amount: 5.6 percent fewer expected mortalities and an expected probability of mortality which is 8.3 percentage points lower, a large effect given the overall mortality rate of 14 percent. A point to note, however, is that the 95 percent confidence interval on the probit marginal effect coefficient is rather broad, ranging from -0.129 to -0.036. Therefore, this effect is not very precisely estimated.

A further point to consider is the fact that the place and person delivering the child only refers to the most recently delivered child. Such controls do not generally appear to explain the expected number or probability of under-five deaths. Environmental factors are far more likely to influence the probability of death of slightly older children. Moreover, some women have had more than one birth in the last two years. Her delivery experiences could be very different for each child but it is not possible to capture this within the models. Therefore, it is plausible that the place and person who delivered the child are relatively unimportant when compared to other factors for such children. What is unexpected is that such factors do not appear to influence the probability of infant deaths; especially when such a substantial proportion of infant deaths are related to birth complications and may therefore be associated with the health services used.

Another consideration is whether or not women access a clinic regularly during their pregnancy. One potential problem with this particular question is what is meant by and what is understood by “regular” attendance at a clinic. It is not clearly specified within the survey. Therefore, what one respondent considers as being regular may not be the same degree of regularity considered by another. As such there is a possibility that this indicator of health services may not be particularly well defined. Nevertheless it is worth examining, as one would expect that attendance at a clinic on a regular basis would be negatively associated with child and infant mortality.

The Poisson results presented in Tables 4.4.1-4.4.5 suggest that attendance at a clinic is consistently negatively associated with the number and probability of under-five and infant deaths. These results are significantly different from zero in all cases. Where a mother attends a clinic on a regular basis, the probability of under-five and infant death all other things being equal is 5 percentage points lower in comparison to a household where the mother does not regularly attend a clinic and this result is statistically significant in all the specifications.[143] Attendance at a clinic may be an indicator for providing mothers with information about how to care for her baby or could be used to identify more risky pregnancies and thus help to assist women in seeking preventative care. Whilst it is not possible to identify exactly what role attendance at a clinic may have and regularity at a clinic is not clearly defined, what is clear is that attendance at a clinic helps to reduce the probability of mortality by a relatively large amount; approximately 5 percentage points. Therefore, it is important to ensure that women have access to such clinics within their local area and to encourage pregnant women to attend such clinics regularly.

4.4.5 Household characteristics

In section 4.3 it was indicated that additional health variables, such as the type of water supply, sanitation and rubbish disposal facilities the household has, would affect households' experience of under-five mortality. Tables 4.4.1-4.4.4 highlight the importance of using mosquito nets. In both the Poisson and the probit models I find that the use of mosquito nets during the mosquito season reduces the expected number of and probability of under-five and infant mortality within Malawi. Table 4.4.1 shows that the use of mosquito nets reduces the expected number of under-five mortalities by just over five percent and this is statistically significant at the 1% significance level.

[143] The level of statistical significance varies by specification. In Table 4.4.1, Table 4.4.3 and Table 4.4.4 attendance at the clinic is significant at the 10% level. In Table 4.4.2 and Table 4.4.5 attendance at the clinic is significant at the 5% level.

Households using mosquito nets are also expected to reduce their probability of experiencing an under-five mortality by seven percentage points. This is a very large reduction indeed. This result suggests that under-five mortality could be halved if households started to use nets for all children aged under-five. The use of bed nets apparently makes a very large contribution in reducing under-five and infant mortality.

Given that 27 percent of all illness related deaths within the sample are associated with malaria,[144] the marginal effect of net use seems rather high. However, one must recall that the cause of death is self-reported and not always confirmed by medical personnel. It is therefore possible that a larger proportion of under-five and infant deaths could be associated with malaria and this may be why the marginal effect of mosquito nets is so very large.

Access to a clean water source and appropriate waste disposal facilities do not generally appear to be significant factors in reducing under-five and infant mortality across households. An exception to this across Tables 4.4.1-4.4.5 is the impact of a household having a flushing toilet. In all cases, such households are expected to experience fewer deaths (of both under-fives and infants) and the probability of experiencing mortality is also substantially lower. In Table 4.4.1 I observe that households which have access to a flushing toilet are 7 percentage points less likely to experience an under-five death than those households which have access to a latrine. This result is intuitively plausible; a flushing toilet ensures that all waste is disposed of quickly and so children are not exposed to human waste which is likely to contain disease. However, it is important to recall that only a very tiny percentage of households have access to a flushing toilet. Therefore, in reality the practical significance of this finding is rather small as so few households have a flushing toilet.

Households with no access to any kind of toilet facilities are found to have a lower probability of childhood death in comparison to a household with a latrine. However, such a finding is not statistically significant in any specification. A possible explanation for this result could be that households with no toilet ensure that they dispose of their waste far from the home and therefore any children are unlikely to come into contact

[144] 17 percent of all deaths within the sample; 99 deaths from malaria and a total of 579 deaths.

with any human waste. The reported marginal effects upon different methods of rubbish disposal appear to have no statistically significant effect upon the probability of experiencing under-five mortality.

4.4.6 Number of births

An examination of Tables 4.4.1-4.4.5 shows that additional births are associated with very large increases in the expected number and probability of experiencing under-five and infant mortality. The results are similar for infants, although the effect is slightly smaller. In all cases the results suggest that this result is highly statistically significant with the effect being far greater than that indicated by any of the other independent variables. The effect is so large that it suggests that an increase in births will actually cause mortality rates to exceed the level of mortality within the sample. This therefore requires some further examination.

As outlined previously, household size and therefore the number of children that women have may be determined by hoarding and replacement behaviour. If households anticipate that they are more likely to experience mortality then they will have more children in readiness for expected losses. Equally households which have previously lost children will seek to have more children as a way of replacing deceased children. The number of births would therefore be expected to be higher in such households.

The number of multiple births is much higher in households where there has been at least one incidence of under-five mortality in the last two years within the sample. It is also clear that the number of mortalities is higher in cases where there are more births. In cases where only one birth has taken place in the last two years, 91 percent of these households experienced no mortality at all. In comparison, if there have been two births, just over a third of these households experienced no mortality. In the very small number of cases where there have been three births, only 5 percent experience no mortality. In addition, if birth spacing is shorter then each birth may potentially have a greater probability of resulting in mortality due to the strain put on the mothers' reproductive system or indeed if the births are multiple births as a result of twins and or triplets. Birth spacing could also be shorter if women give birth prematurely and subsequently go on to conceive again very quickly (particularly in the case of the first child dying). As also noted within *Chapter 2*, if children are born in relatively quick succession, then exclusive breastfeeding is likely to be shorter, which may mean that existing young children are no longer breastfed. Such children may then face potential exposure to disease from alternative feeding methods or reduced immunity which is derived from the mother. It is therefore possible that the very large effect of additional births is being driven by the fact that these births would appear to be more likely to result in mortality. In anticipating or having experienced mortality such women have additional children in relatively quick succession which in itself then reinforces the probability of mortality due to reasons outlined above.

4.4.7 Extensions to the model

On the whole the overall fit of the models presented for children under-fives are similar to those presented for infants. Table 4.4.1 indicates that the Poisson model Pseudo R-squared is in the region of 0.18 and 0.22 for the probit model. In comparison in Table 4.4.2, the reported pseudo R^2 for infants using the same specification as used in Table 4.4.1 is 0.19 for the Poisson and 0.20 for the probit results.

As I noted earlier in this chapter, households report the cause of death in the majority of cases. Within the literature, there does not appear to be any differentiation between the causes of death. Therefore, this section presents a new line of exploration.

Using the estimation presented in Table 4.4.4, the cause of death is now taken into account by examining illness and non-illness related deaths. In line with the discussion presented in section 4.3.1, a set of new dummy variables are created which are coded according to whether the household experienced the different types of mortality. These categories are; under-five illness related death, under-five non-illness related death, infant illness related death and infant non-illness related death.

Table 4.4.5 Probit marginal effects with under-five and infant deaths illness and non-illness related deaths as the dependent variable.

	Under-five		Infant	
	Illness related	Non-illness related	Illness related	Non-illness related
logy	0.002 (0.003)	-0.003 (0.002)	0.001 (0.002)	-0.003 (0.002)
logydummy	0.010 (0.028)	-0.028 (0.014)**	0.010 (0.021)	-0.024 (0.012)**
north	-0.028 (0.011)***	-0.017 (0.006)***	-0.017 (0.006)***	-0.014 (0.005)***
centre	0.017 (0.010)*	-0.017 (0.006)***	-0.005 (0.006)	-0.018 (0.005)***
rural	-0.008 (0.015)	0.025 (0.007)***	-0.010 (0.008)	0.025 (0.005)***
educated	-0.035 (0.011)***	-0.011 (0.008)	-0.017 (0.006)***	-0.003 (0.008)
educdummy	-0.001 (0.010)	-0.014 (0.006)**	-0.011 (0.005)**	-0.012 (0.005)**
employed	-0.011 (0.010)	-0.008 (0.007)	-0.0003 (0.006)	-0.006 (0.006)
goclinic	0.001 (0.017)	-0.055 (0.022)***	-0.002 (0.011)	-0.047 (0.020)**
clinic	0.026 (0.015)*	-0.017 (0.007)***	0.006 (0.009)	-0.016 (0.006)***
at home	0.040 (0.037)	0.016 (0.020)	-0.007 (0.013)	0.001 (0.014)
other	0.089 (0.107)	dropped	0.038 (0.053)	dropped
attend	0.107 (0.061)*	0.019 (0.028)	0.003 (0.018)	0.0002 (0.017)
on way	-0.009 (0.055)	0.014 (0.043)	-0.014 (0.021)	0.017 (0.040)
doctor	0.009 (0.018)	0.028 (0.017)*	0.004 (0.012)	0.033 (0.017)**
midwife	0.007 (0.019)	-0.010 (0.010)	0.010 (0.013)	-0.007 (0.009)
TBA	-0.059 (0.019)***	-0.026 (0.011)**	-0.0008 (0.014)	-0.011 (0.012)
self	0.008 (0.029)	-0.004 (0.017)***	0.024 (0.200)	0.003 (0.016)
births	0.232 (0.015)***	0.107 (0.010)***	0.095 (0.009)***	0.074 (0.008)***
allnet	-0.043 (0.009)***	-0.024 (0.006)***	-0.029 (0.005)***	-0.027 (0.005)***
somenet	-0.044 (0.013)***	-0.029 (0.006)***	-0.022 (0.006)***	-0.023 (0.006)***
pipewater	0.008 (0.014)	0.005 (0.010)	0.019 (0.010)*	0.008 (0.009)
wellspringres	-0.003 (0.009)	0.002 (0.006)	-0.005 (0.006)	0.001 (0.023)
flush	-0.053 (0.012)***	-0.002 (0.022)	-0.022 (0.007)***	0.008 (0.023)*
notoilet	0.0001 (0.011)	-0.012 (0.007)*	0.002 (0.007)	-0.011 (0.006)*
bin	0.008 (0.035)	-0.025 (0.010)***	0.030 (0.031)	-0.016 (0.012)
heap	-0.007 (0.011)	-0.003 (0.008)	-0.002 (0.007)	-0.002 (0.007)
norubbish	-0.002 (0.011)	0.010 (0.008)	0.001 (0.007)	0.012 (0.007)
N	3597	3410	3607	3572

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses.

The alternative category for all cases is that no[145] death has occurred (at all). As a result of this, the number of observations in each estimation falls by the number of excluded cases.[146] The estimation results using illness and non-illness related deaths as the dependent variable indicate some similarities and differences when compared to the results presented previously.

A closer examination of the marginal effects reported in Table 4.4.5 suggests that there are differences in the factors which influence the probability of experiencing under-five or infant mortality given different causes of death.

Particular variables of interest worth noting are rural residence, whether or not the mother is educated, whether the mother attended a clinic during pregnancy on a regular basis and the number of births that occurred in the last two years. In Table 4.4.4 I observed that there was no statistically significant impact on under-five or indeed infant mortality due to living in the rural area. The results presented in Table 4.4.5 however, suggest that rural residence does have an impact on the expected probability of non-illness related deaths, albeit a modest increase of 2.5 percentage points as previously noted in section 4.4.3 where it is speculated that rural residence could be associated with other factors such as the quality of health services.

Moreover, the role of education in terms of reducing the probability of illness related deaths is reasonably large at 3.5 percentage points for under-fives but falls to just over one percentage point for non-illness related deaths. As noted in section 4.4.2, such a finding could be attributed to the fact that non-illness related deaths are commonly due to birthing complications. It is very unlikely that the mothers' level of education would have any influence on the probability of death in such cases as sadly such matters are primarily outwith her control. However, a mother who has some education may be more able to make decisions which lead to better illness prevention and so in turn reduce the probability of illness-related deaths.

Regularly attending a clinic during pregnancy reduces the probability of experiencing a non-illness related death and this is statistically significant at the 5% significance level

[145] No under-five or infant death.
[146] Non-illness related deaths are not included in any estimation of illness related death and vice versa.

but this is not the case for illness related deaths. Women who have attended a clinic with some degree of regularity may be monitored and advised on how to care for their baby. Given the large number of deaths which arise as a result of birth complications, it is possible that regular clinic attendance helps to identify such women and ensure that they are properly monitored and supported to reduce such risks.

In Table 4.4.1 it was observed that an increase in the number of births experienced by each mother increases the expected probability of mortality for under-fives by 32 percentage points This is clearly a huge effect and does not seem plausible given the overall rate of mortality. In Table 4.4.5 where the cause of death is taken into consideration the effect of additional births is still very high but drops to 23 percentage points for illness related deaths. As I have previously suggested, it is possible that the very large marginal effects coefficient is being driven by a relatively small number of women who have experienced multiple mortalities. A more ideal measure would have been information on the exact number of births each mother had experienced within the last 5 years.

4.4.8 Summary of findings

Income is not found to be statistically significant in explaining the number of and probability of under-five and infant deaths in Malawi. I tested whether this finding was perhaps sensitive to my definition of income and so I also ran regressions using different measures of household income such as net income from agriculture during the dry and rainy season. Using alternative measures of income does not alter the results. Therefore, I conclude that I can find no evidence to support the proposition that, ceteris paribus (including holding education constant), increased household income will significantly reduce under-five and infant mortality in Malawi.

However, it should be noted that the measures of income that I am using are recent household incomes, generated in the last 12 months or in the last cropping season. One might reasonably expect that as health (or lack thereof) is a stock; it would be past income rather than current income might be a more appropriate measure to consider, in particular, household income over a number of years. Moreover, the relationship (if one is to believe that there is one) between income and death is also likely to be weaker when deaths have occurred less recently. Unless incomes were fairly constant over time, and therefore current income was also reflective of past income, the role of current income is unlikely to have any influence on past (more than one year ago) deaths. Ideally longitudinal data would be preferential so that I could explore this line of investigation within future research.

Even if a more robust measure of household income could be used within the study, there is also another possibility that should be considered given that the overall effect of household income upon under-five and infant mortality is tiny. Grossman (1999) argues that education is a *"more important correlate of health than occupation or income, the two other components of socioeconomic status"*[147] and that education is a strong determinant of income. Taking the Grossman (1999) argument, if education is a relatively strong correlate of socioeconomic status (and income) it could be that the impact of income is small because (as a socioeconomic determinant) it is in fact being captured by the coefficient on education. Therefore, considering the Mosley and Chen (1984) framework, the key path through which the socioeconomic determinants work their way into influencing the proximate determinants is not via income but actually via education which is effectively encapsulating socioeconomic status.

A number of other independent variables that I thought would be potential determinants of determining child and infant mortality also do not appear to have any significant impact. Age and age-squared were not statistically significant in any of the specifications and therefore these two variables were dropped in my models and do not appear within the presented results. Labour force participation was also generally not found to be associated with child mortality.

I also find that households' perception of food consumption and health services are not statistically significant determinants of under-five or infant mortality in Malawi. These particular indicators are not specifically targeted at the mother or at young children. Therefore this could explain why no relationship is found. Equally of course it is

[147] p.65.

possible that even if I had information on these variables specifically for my two parties of interest; mothers and children, there may not actually be a relationship to observe.

Other health indicators such as water source, type of sanitation and rubbish disposal are also found not to have any significant impact on child mortality conclusively. The findings related to who assisted the mother to give birth are mixed. Women who give birth without the assistance of trained medical personnel tend to experience more mortalities and a greater probability of mortality than those who are assisted by a nurse. Moreover, there is some limited evidence to suggest that women aided by a doctor seem to be at a greater likelihood of experiencing mortality than a woman assisted by a nurse, all other things being equal. As outlined in section 4.4.4 this could possibly be attributed to women seeking assistance from more specialised medical personnel when they have more complicated pregnancies and the birth of their child is more likely to result in mortality.

A key finding that becomes very apparent in the results is the importance of the use of mosquito nets in reducing the number of and probability of under-five and infant mortality. Given that malaria is the leading cause of illness related deaths in Malawi the results reconfirm the importance of using preventative measures against malaria.

Finally, the number of births experienced in the last two years is highly significant in all specifications. Women who have experienced mortality tend to have more births relative to those who have not experienced mortality in the sample. This could be because women seek to replace such children or in anticipation of losses have more children in readiness for such losses. A higher number of births within a short time period can be associated with either short birth spacing or multiple births. Very short birth intervals could be associated with pregnancies that were not to full term, with premature babies being more likely to result in mortality than a full term baby. Pregnancies which are multiple births such as in the cases of twins may also be more risky and likely to result in complications and potential deaths. These factors could be what is driving the very large effect of births: births in quick succession due to previous losses and the close proximity of each birth in turn, increase the probability of death in future.

One final issue which is worth reiterating is the potential for measurement error for the births indicator. The household roster and deaths module lists children (collectively) as both biological and adopted children within the household. Therefore the calculation of the number of births per mother may be overestimated, if there are some adopted children within each household. There are however no means to separate biological and potentially adopted children from one another within the IHS2 to gain a better estimate of the number of births. It is therefore not possible to be confident about the magnitude of the additional births as there may be potential measurement error.

4.5 Conclusion

The aim of this chapter was to investigate the determinants of child and infant mortality in Malawi. In order to provide the best possible opportunity to reduce child mortality, there needs to be a clear understanding of the determinants of child mortality in different environments.

Education as in previous studies appears to be an important factor in reducing mortality at the primary level. If we consider education more generally, I observe that possessing at least some education does have a positive impact in reducing the expected number of and probability of child mortality. As Basu and Stephenson (2005) also found, even at very low levels of education, the impact in terms of reductions in child mortality can be quite substantial. This provides further evidence to support the continued expansion of the provision of primary education as a very minimum. Manda (1999) also finds evidence that in Malawi education only has what he refers to as a *"modest impact"*[148] on child mortality. A suggestion put forward is that as only a very small proportion of women in Malawi possess education beyond primary level; *"appreciable effects"*[149] of education may not exist. The results presented within this chapter support this proposition within the Malawian context.

Manda (1999) suggests that for any real gains to be achieved education levels within the country would need to be raised quite significantly and this would take a considerable

[148] p.310.
[149] p.310.

amount of time. Given that there is evidence supporting the positive impact primary education can have and relatively speaking the data indicates that rates of primary education in Malawi are low, this should be one of the first priorities in terms of expanding education. Longer term, there are opportunities to build upon this, once primary level education is more universally accessible.

Some of the health services related findings indicate a need for additional information about the mother, her pregnancy and her decision regarding where she gave birth and who assisted her in delivery. The DHS provides more information about each and every single birth which would be more beneficial as the data presented within the IHS2 only considers the most recent birth and experiences could differ in each case.

Perhaps one of the most important findings from this chapter is the significant role of mosquito nets in reducing child and infant mortality. Where nets are available and provided to all children under the age of five, the reduction in the number of and probability of experiencing a under-five or infant death is not to be ignored and provides evidence in support a more widely spread programme for universal use of bed nets.

As I noted at the end of the last section, the number of children born to each mother has a very large and statistically significant affect upon under-five and infant mortality in all of the specifications. An explanation proposed for such a result is that a small number of women who experienced multiple births and subsequently a greater incidence of mortality are driving this result. The overall mortality rate within the sample is 14 percent whereas the marginal effect on births is over 30 percentage points in Table 4.4.4. Moreover, the inclusion of birth spacing information would have been useful to examine whether the order in which children are born has any effect upon their probability of mortality. However, a lack of data makes it impossible to make any comment on these issues.

In conclusion, the findings presented in this chapter provide some supporting evidence for some of the commonly thought of factors which are positively or negatively associated with child and infant mortality. This chapter also shows that there a number of factors which do not appear to be associated with child and infant mortality in either the direction one would expect and/or in a statistically significant manner. Therefore further research is required to improve our understanding of the factors which shape child and infant mortality patterns in Malawi.

Chapter 5: The impact of morbidity on wages in Malawi

In June 2006 the World Bank published a poverty and vulnerability assessment related to the inhabitants of Malawi. The report found that *"chronic illnesses and HIV/AIDS, are major challenges playing an overwhelming role in the daily life of Malawians."*[150] Morbidity may be problematic if it leads to withdrawal from the labour market or on-farm activities; resulting in a loss of wages or productive activity. In response to morbidity individuals may seek medical care which causes a rise in healthcare expenditures and as noted in *Chapter 2,* such expenditures can be large. Indeed a report by the World Bank (2006) found that 14 percent of Malawian households experienced healthcare expenditures which amounted to a quarter of their annual income. Large expenditures together with a loss of wage income can potentially have longitudinal effects on household wealth and future welfare.

This chapter examines the impact of morbidity upon wages (as a source of earned income) within Malawi using the human capital literature developed by Shultz (1961), Ben-Porath (1967), Mincer (1974) and Becker (1962).

Whilst there are other studies on this topic for a number of other developing countries as reviewed in *Chapter 2,* none of them have examined Malawi.[151] Malawi is an interesting case as it has a relatively high rate of HIV prevalence at 10.3 percent compared to 4.5 percent for sub-Saharan Africa as a whole.[152] It also has a high incidence of malaria and is classed as a high transmission country. Such potentially serious health issues therefore present an interesting case in a primarily rural setting, in contrast to existing studies which have primarily included urban waged workers.

In the standardised earnings estimations such as those developed by Mincer (1974) the number of years of schooling is included to provide an estimate of the return to education. Education, however, is not the only investment in human capital that one can make. Human capital also includes: health, on-the job-training, occupational choice and migration (Shultz, 1961; Ben-Porath 1967). By undertaking such investments

---

[150] p. xii

[151] One study specifically related to HIV/AIDS has been undertaken for Malawi by Dorward and Mwale (2005) but this did not consider morbidity more generally.

[152] http://wdi.worldbank.org/table/2.20

individuals seek to receive returns in the form of higher wages and earning capabilities over their lifetime (Grossman, 1972; Ben-Porath, 1967; Shultz, 1961; Mincer 1975).

This chapter seeks to incorporate a proxy for health (effectively a lack thereof) known as "disabled days" into a traditional wage estimation to assess the impact of morbidity on Malawian wages. The specific model used within this chapter follows an approach developed by Schultz and Tansel (1997) using the Integrated Household Survey (IHS2) which is outlined fully in *Chapter 3.*

The remainder of this chapter proceeds as follows: *Section 5.1* outlines the model framework and research question. *Section 5.2* presents the variables that will be included within the estimations and some descriptive statistics. *Section 5.3* highlights the key results from the regression estimates and *Section 5.4* concludes the chapter.

5.1 Human capital and wage earnings

Becker (1962) states that activities which we undertake to influence future incomes via the *"imbedding of resources in people"* is investment in human capital. Fundamentally, Shultz (1961) argues that people undertake such investments as it has an economic value within the labour market. People who have undertaken such investment; whether it is in on-the-job training or better health possess greater physical and mental ability relative to those who have not made such investments. These investments are costly in terms of time, forgone earnings and direct expenditures (Grossman, 1972). Therefore one would only make such investments if there was an appropriate payoff. The payoff is that firms value such investments, as people are more productive and therefore raise income prospects (Becker, 1962).

The Mincer (1974) earnings equation, which is briefly outlined in *Chapter 2,* relates investments in human capital to the observed differences in individual wages or earnings. The standard Mincer (1974) earnings function takes the log of wages and includes regressors such as the number of years of schooling, experience and the square of experience to capture the rise and then decline in earnings over time. In practice, experience may not be directly observed and therefore a proxy of experience may be used. This is calculated by Mincer (1974) as the difference between an individual's age and their years of schooling; assuming they began school at age six and started work immediately after leaving school:[153]

$$\ln y = \beta_0 + \beta_1 s + \beta_2 x + \beta_3 x^2 + u$$

Willis (1986) and Becker (2007) indicate that the vast majority of the literature examining human capital has concentrated on the role of education and training, which is clearly part of an individual's collective human capital (Becker, 1962; Becker 2007; Grossman, 1972, Shultz, 1961). The approach used within this chapter will use a Mincer-type wage equation and add an indicator for health as proxy for an additional indicator of an individual's human capital. The results presented within this chapter can be viewed alongside existing estimations of the returns to human capital in the traditional sense of education and experience.

The measure of health used within this chapter is a measure of morbidity which will be referred to as disabled days, closely following the work of Schultz and Tansel (1997), and will seek to estimate the effect that morbidity has on individuals' wage rate and other earned-income received. If morbidity leads to small reductions in wages then it may be possible that the collective household unit could make adjustments to compensate for the reduction in household income. However, if the reduction in wages is large and/or illness is widespread, meaning that many individuals experience morbidity, the ability of the household to fully compensate for such reductions may be limited.

This chapter will utilise the 2004 Malawi Integrated Household Survey (IHS2) which is a nationally representative sample survey to capture information on a range of different aspects associated with household welfare in Malawi which was undertaken throughout 2004-2005 by the National Statistical Office. A review of the survey data is detailed in *Chapter 3*.

[153] $x = a - s - 6$ (where x is experience, a is age in years and s is years of schooling).

5.2 Variables used for regression estimates

In this section, the key variables which are subsequently used within the regression estimations are outlined. Health has been described as *"a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity"*.[154] With that in mind, measuring health status is not straightforward as it is a relatively broad concept and is also subjective. What one person considers to be "good" health may not be the same for another. Moreover, when talking about measuring health, one needs to be clear about specifically what aspect of health one is measuring. A full discussion of the issues related to the measurement of health status can be found in *Chapter 2*.

A key data module within the IHS2 is the health module which collected information about the general health of each household member. The measure of morbidity which is used within this chapter is referred to as "disabled days" in line with Shultz and Tansel (1997); it uses three specific questions from the health module of the survey.

In the first stage, respondents were asked *"during the past two weeks have you suffered from an illness or injury?"* This is used to determine whether or not a household member has experienced any recent morbidity prior to the IHS2 survey being undertaken. Only those individuals who indicated that they have experienced illness or injury answer the next question in the second stage. A follow up question is then used; *"during the past two weeks did you have to stop your normal activities because of this (these) illness(es)?"* to determine whether someone has stopped their activities due to any illness or injury that they have experienced. Together these two questions provide a measure of morbidity in that they determine whether or not someone was disabled due to their condition and are shown as the first two steps in Figure 5.2.1.

Someone may be ill but they may not have stopped their normal activities. For an individual who has stopped their normal activities there is then a third stage of questioning. This determines the length of disability which will subsequently be referred to as disabled days. The third question is "*for how many days in the past two*

[154] http://www.who.int/about/definition/en/print.html

weeks did you have to stop your normal activities?" This final part of the selection process then determines a measure of morbidity as shown in the final panel of the diagram below. Therefore, the measure of morbidity used refers to someone who experiences disabled days; this is someone who has both been ill/injured in the two weeks prior to the IHS2 being undertaken *and* has stopped their normal activities for at least one day.

The use of a measure such as disabled days has a number of characteristics which warrant consideration. Individuals self-report their health status within the two-week period; indicating whether they had an illness or injury and whether that illness or injury prevented normal activities from taking place. This enables the respondent to make a judgement about the severity of their condition: whether it is sufficiently debilitating to stop them from working. There are however some potential reporting errors which can arise with such an approach and one should be aware of.

The first is that of mis-reporting, due to respondents having to recall their recent experiences. People may indicate that they were “disabled” whereas in fact they were not or vice versa. Given that the recall period is fairly short and also fairly recent, the degree of mis-reporting may not be hugely problematic. If, however, the recall period were less recent or over a longer time period, there would be more of a concern regarding the accuracy of the responses provided.

A second problem previously is that certain groups within society such as the wealthy and more educated may report incidences of illness more readily than those of lower income and they may also overestimate the impact that illness has on their health relative to the actual clinical impact (Shultz, 2005).

This particular problem may be viewed as one of “cultural conditioning” and this behaviour may go some way in explaining the presence of increasing morbidity rates alongside increases in life expectancy (Shultz and Tansel, 1997). Their hypothesis is that we all have a threshold of what we perceive to be good health and that within different social and cultural settings this threshold may differ quite substantially. Hence, when using self-reported estimates of morbidity it is important to be aware of this potential over estimation issue (Johansson, 1991; Shultz, 2005).

The third problem is that accurately measuring morbidity for self-employed individuals is difficult. Self-employed individuals are likely to reduce their hours or effort which would mask the impact of illness on wages and thus we underestimate the impact of morbidity (Shultz and Tansel, 1997). Therefore, Shultz and Tansel (1997) put forward a case for using only waged workers. In contrast, to self-employed individuals, waged workers are argued to report morbidity more accurately (with a lower level of bias), as a reduction in their activity is more readily observed by fellow colleagues (if, as Shultz and Tansel (1997) assume, workers are interdependent). Therefore, workers will only claim to have been sick if they really are sick, as there is no gain to be made from shirking and employers will wish to minimise shirking as it will reduce production.

Another issue to note is that the measure of disabled days is specific to a reference period. It can be argued that a self-reported measure is broader and more inclusive measure of *current* health compared to that which may be officially diagnosed stated by medical professionals (Idler and Benyamini, 1997; Blane et al., 1996) which makes self-reported health measures very attractive. However, self-reported measures may only capture morbidity at a very specific point in time if a reference period is given, which may not necessarily be reflective of morbidity over time and individual health status more generally in the case of acute health conditions. Equally however, self-reported measures could be reflective over a period of time; typically a year.[155] Therefore, when using self-reported measures, the reference period is important to consider as it can influence the response. For example, Idler and Benyamini (1997) demonstrate a scenario where the respondent is asked to rate their health on a scale. This effectively enables the individual to take into account different aspects of their health in order to determine their overall health status and incorporate a time dimension into their rating of health; such as a comparison to their health a year ago; such information is also collected within the IHS2 and is considered within the estimations which follow. The disabled days measure used within this chapter is much more specific and the respondent is therefore constrained to a particular period.

[155] A number of studies considered in Chapter 2 indicate that people evaluate their level of health on a scale and this may be in comparison to their health a year previously.

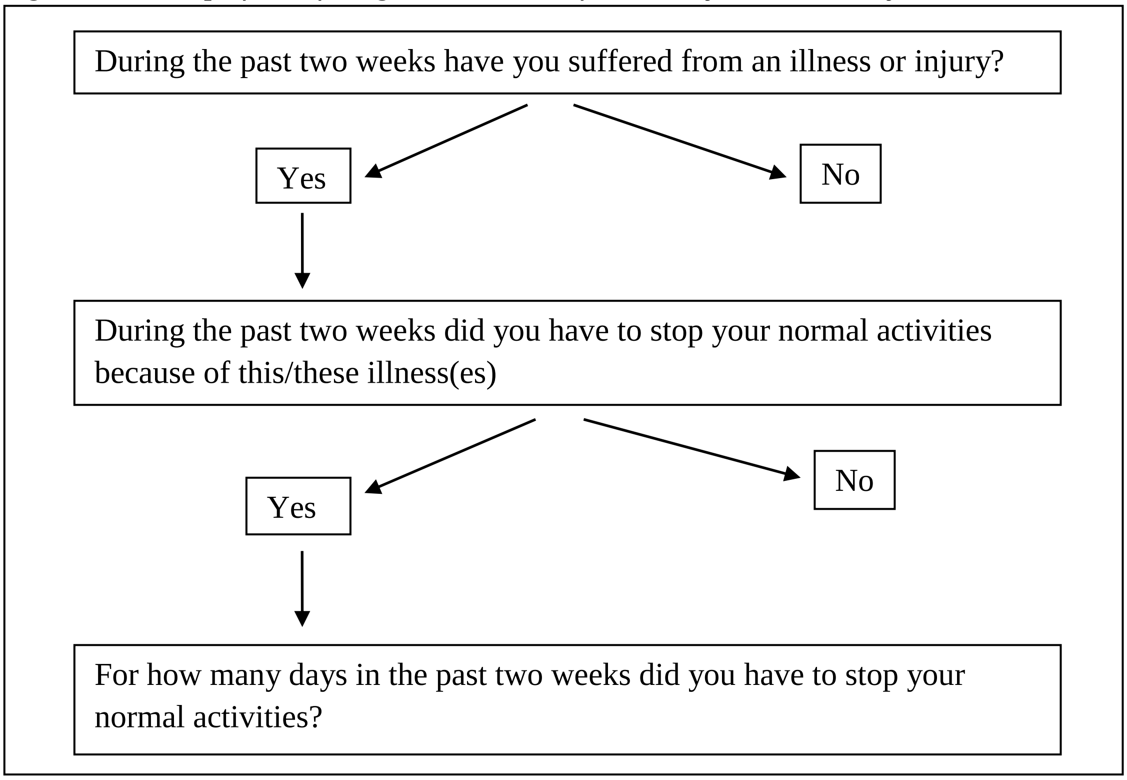
In comparison to the Schultz and Tansel (1997) study, the IHS2 data only captures information regarding morbidity over a two week period. This is a relatively short period of time within a year and may fail to capture acute morbidity for a number of individuals given the short time period being considered; for example, an individual having a broken leg. However, such health ailments are not particularly relevant in terms of the human capital theory being explored within this chapter. Whilst a broken leg would reduce an individual's stock of human capital in the short run, in the long run they should (assuming no complications) be able to restore their stock of capital and one would expect that the impact on their wage would be negligible. This would be in contrast to someone who had a more serious long-term health condition, where such an individual may find it much more difficult to replenish their stock of human capital. This may be particularly true in cases where they experience multiple periods of morbidity which would cause further erosion of their stock of human capital.

A shorter reference period may exclude the former acute periods of morbidity and in so doing then the effect of morbidity would be underestimated. However, the latter group is really the focus and it could be argued that if someone is generally experiencing morbidity (rather than an acute condition), then there is a reasonable probability that they will be "disabled" within the two week period. Therefore, the impact of morbidity may be more accurately estimated by using a smaller reference period. Equally, there may be some individuals who have experienced a period of acute illness or injury within the two week period which is not reflective of their general health and stock of human capital but who would be included within an estimation of the impact of morbidity on wages. In such circumstances, one would overestimate the impact of morbidity.

There is no reason to believe that those experiencing acute periods of morbidity are any more likely to be included than not included within the two week reference period. The two week period is randomly distributed across the entire sample based on when enumerators were within an area. Therefore, it would appear that the inclusion or exclusion of individuals experiencing acute morbidity would not appear to be biased in one direction rather than the other. The main difference to the Schultz and Tansel (1997) study is that the maximum morbidity period is two weeks rather than four and therefore the results are not directly comparable. It should also be noted that there may be individuals who experienced morbidity beyond the two week period, but it is not possible to identify such individuals.

In addition, to the three key questions just outlined and indicated in Figure 5.2.1, individuals were also asked what type of illness or injury they experienced, who diagnosed the condition, actions that were undertaken and whether anyone else stopped their normal activities to provide care. As the IHS2 is also part of the LSMS which is outlined in *Chapter 3*, the questions posed in this survey are almost identical to those used in the study of Ghana and Cote d'Ivoire by Shultz and Tansel (1997) and therefore, a similar approach is undertaken and some comparison of the results can be considered.

Figure 5.2.1 Steps for defining the measure of morbidity: disabled days



Source: IHS2

Shultz and Tansel (1997) highlight that the use of waged employees means that the impact of morbidity on wages is only estimated for those individuals. This approach excludes a significant proportion of individuals who engage in family-based operations or those who operate as individual home production workers. In low-income countries a large proportion of the population may be engaged in such activities and as such the estimates presented will not be representative of the impact of morbidity on incomes for the population as a whole.

Effectively only a sub-sample of the entire population is observed within the estimations presented, in this case waged/ganyu individuals rather than the whole working population, as rates can only be identified for those who have worked in the last 12 months. Therefore the estimate obtained for the effect of disabled days on wages/ganyu rates applies only to those who currently report a wage or ganyu rate and not to the impact of morbidity on individual wages more generally. It is possible that morbidity could have been so severe that some individuals could leave the labour force all together. The exclusion of such individuals would lead to an underestimation of the effect of morbidity on such daily rates.

Another issue worth considering is the proportion of individuals who participate in the wage sector relative to those who do not based on the information provided within the IHS2. If only a small fraction of the population is employed in waged activities, then the results obtained from the regression analysis will be informative only with respect to a small part of the economy.

Within the IHS2, fourteen percent of 22,998 prime aged individuals report a formal wage. Prime age individuals are defined as those aged between 14 and 49 years of age.[156] Slightly more than a fifth (22%) of prime aged males participated in waged employment, whilst only 6.5% of the comparative female group are engaged in waged employment.

The participation rates in waged employment are relatively similar to that found by Shultz and Tansel (1997). In both cases a relatively small number of females are engaged in waged activities whereas approximately one in four to one in five males participate in waged activity. Shultz and Tansel (1997) only consider male workers and report their findings for women in a separate paper Shultz and Tansel (1992).

[156] Inclusive.

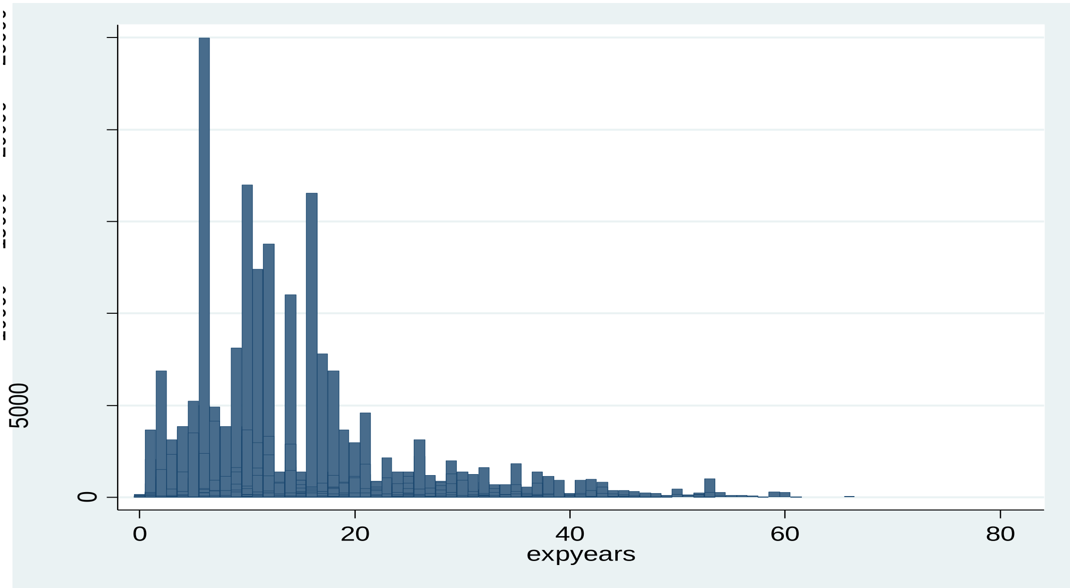
In order to examine the effect that morbidity has on an individual's wage, information about the daily wage received is also required. Respondents to the IHS2 were asked whether or not *"At any point in the past 12 months, were you employed for a salary or commission or any payment in-kind excluding ganyu?"* This question identifies all individuals who have received some form of waged income in the last year which then forms the sub-sample for analysis.

Individuals are also asked what their last payment was and the period of time that the payment covered; this may be reported in days, weeks or months. From this an estimated daily wage for each individual can be calculated. In addition to the formal wage measure outlined above, respondents were also asked to indicate the value of any in-kind payments they received in a similar manner as well as whether they participated in any ganyu labour. The reported wage from ganyu labour was reported as a daily amount and therefore no further calculation was required for this indicator.

Additional control variables are used in the model in line with those used in the human capital theory literature to capture socio-demographic characteristics such as gender, residence, level of education and experience (Beegle, 2003; Shultz, 2005; Thomas and Strauss, 1997; Shultz and Tansel, 1997) as outlined in additional detail in *Chapter 2*. As observed in *Chapter 3*, individuals in the IHS2 report education in both levels and the number of years of education that they had completed. Years of education is measured with a large amount of error in some cases. Therefore a completed level of education is preferable for inclusion in the regression estimations. Each level of education is included as a dummy variable with the exclusion of no education as the base category for comparison. As was the case in *Chapter 4*, individuals who had missing information for education are included by the provision of a dummy variable labelled as educdummy. For these individuals it is not clear whether the level of education is truly missing or is in fact indicative of someone not having achieved any level of formal education (a category which was also possible). Therefore the approach given is to include a dummy to ensure their inclusion without making any assumptions about their education level.

Experience and its square are also included as standard variables in a Mincer (1974) type earnings equation. As anticipated, wages are initially rising for those at the bottom of the age spectrum and then decline once an individual heads towards retirement. This pattern is reflective of the fact that the stock of collective human capital (which is inclusive of health) that an individual possesses over their lifetime is correlated with age (Grossman, 1972). Figure 5.2.1 shows the experience-wage profile for Malawi using the IHS2. To capture the inverse-U pattern, both experience and the square of experience are included in the regression estimation. Experience has been calculated by taking the difference between the year that the individual left school and the year which the survey took place.

Figure 5.2.2 The experience-wage profile for Malawi



Source: IHS2[157]

The area of residence has also been included within the estimations, as wages are likely to be influenced by the location of the job. The urban area attracts migrants as wages are higher than those in the rural area and so residence status is an important control variable. In the IHS2, residency status is reported in a number of different ways. Location based on geographical region within Malawi. Northern and Central dummies have been included and the Southern region has been excluded as the base category.

[157] Wages are measured in Malawian Kwacha. To give an indication of the Malawian Kwacha value in UK Sterling, 195 MK was approximately £1 over the survey period.

Alternatively, a slightly more detailed variation of the rural:urban split is also reported, breaking residency down into an additional five categories.[158]

In addition to the controls outlined above, Shultz and Tansel (1997) also include controls for unobserved human capital transfers that are proxied by parent education and occupation. One would anticipate that parents who are well educated would themselves be in a better paid job ceteris paribus. Hence such parents may be more able to make investments in their children which would lead to gains in their human capital in adulthood.

It is not possible to add in these controls to the model as directly tractable information on each individuals' parents is not readily identifiable. The only possible way to extract this information would be to identify the person code of each working individual and then attempt to identify their parent (mother, father or both) and then extract information that they report. As noted in *Chapter 4*, this process is not straightforward and each individual is linked to the head of the household rather than by their biological or marital relationships, and therefore this has not been explored within this thesis.

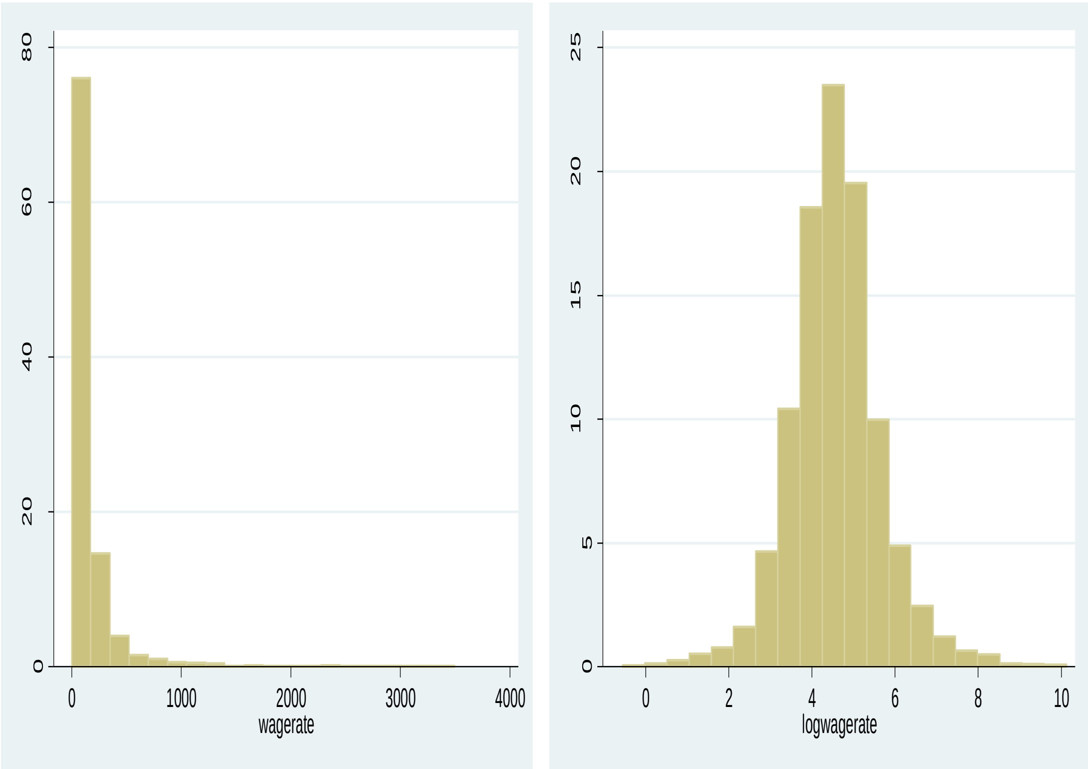
5.2.1 Model framework and considerations

The underlying estimation approach used is based on the Mincer (1974) wage equation developed within the human capital theory literature, where differences in wages are explained by differing endowments of human capital amongst individuals such as education and experience. An additional indicator of human capital is also included within the model to take into account of the fact that an individual's stock of health also forms part of their overall stock of human capital. Within the Grossman (1972) model, investments in health enable an individual to reduce periods of morbidity which leads to less time in the workplace. Therefore, in monetary terms this investment is realised as a wage as the individual can continue working when they are in good health (Grossman, 1999).

[158] The additional breakdown into five categories is not utilised within the estimations presented.

The measure of health included within the estimations presented is proxied by the incidence of morbidity, an approach used by Schultz and Tansel (1997). In particular, this refers to morbidity that has been experienced by the wage-earning individual within the two week reference period prior to the survey. The model will initially be estimated using OLS in line with the standard Mincer (1974) wage equation. In addition some controls are included as a series of dummies for the location of the individual and the type of employer that the individual had at the time of the survey.

Figure 5.2.3 Distribution of wage rate and log wage rate



Source: IHS2

The use of a log transformation for the dependent variable has two properties which are desirable. The first is that it takes into account that there may be extremes within the data; there could be a small number of individuals who report very high levels of wages whereas in fact the rest of the sample have much more modest wages. By taking a log transformation, such extreme values are compressed and more normally distributed as shown in Figure 5.2.3.[159] The other benefit of such a transformation is that it enables the coefficients on the independent variables to be interpreted as semi-elasticities.

A problem with this choice of functional form is that for a number of dependent variables such as; wages, wealth or income, the reported value often is zero. People may not have any wealth as they spend everything they have and don't save, or have no income as they receive no pension or earnings from other activities such as property rental, or it could be that they have no wage as their activities are all on farm based and they are not employed in the traditional sense. As a result of this such individuals would report a value of zero for these characteristics. The problem is that log of zero is not well defined; therefore, anyone not reporting a wage for the purpose of this chapter is not included within the estimations reported.

In order to deal with such a problem, Burbridge et al. (1988) consider the use of the inverse hyperbolic sine transformation[160] but also note that another potential consideration which is used by Diamond and Hausman (1984) cited within Burbridge et al. (1988) is to recode the data with a value of one for such cases to ensure that the observation is included. The inverse hyperbolic sine transformation[161] ensures that values which were initially zero can be included (so the full sample can be analysed) and the coefficients can be interpreted as if they were a logarithmic transformation (except when the value of the dependent variable is very small). However, within the IHS2, it is not possible to identify whether everyone else would be potentially employable and therefore part of the labour force which would be relevant for inclusion within the estimations. Hence, in the absence of information which can identify such individuals; utilisation of the inverse hyperbolic sine transformation is redundant.

As reviewed in *Chapter 2*, people make choices about how much they invest in their health and other forms of human capital depending on the costs of these investments and the expected return resulting from improvements in productivity (Grossman, 1972;

---

[159] The wage rate histogram presented includes 99% of all observations. Less than 1% of all respondents indicated that they received an average daily wage rate in excess of 3500 MK.

[160] The other proposal outlined within their paper is the Box-Cox transformation but this is not applicable to cases where the value is zero. Therefore, the Box-Cox transformation is not considered in any more detail.

[161] The transformation is given by: $\log(y_i+(y_i^2+1)^{1/2})$

Shultz, 1961; Ben-Porath, 1967; Becker, 1962). Individuals who are in better health are more able to work and will be more productive, leading to higher wages. In turn such individuals make greater investments in their health via the consumption of medicine, care and other related activities such as better nutrition which is afforded to them due to their higher wage (Grossman, 1972; Grossman, 1999; Lee, 1982; Shultz, 2005).

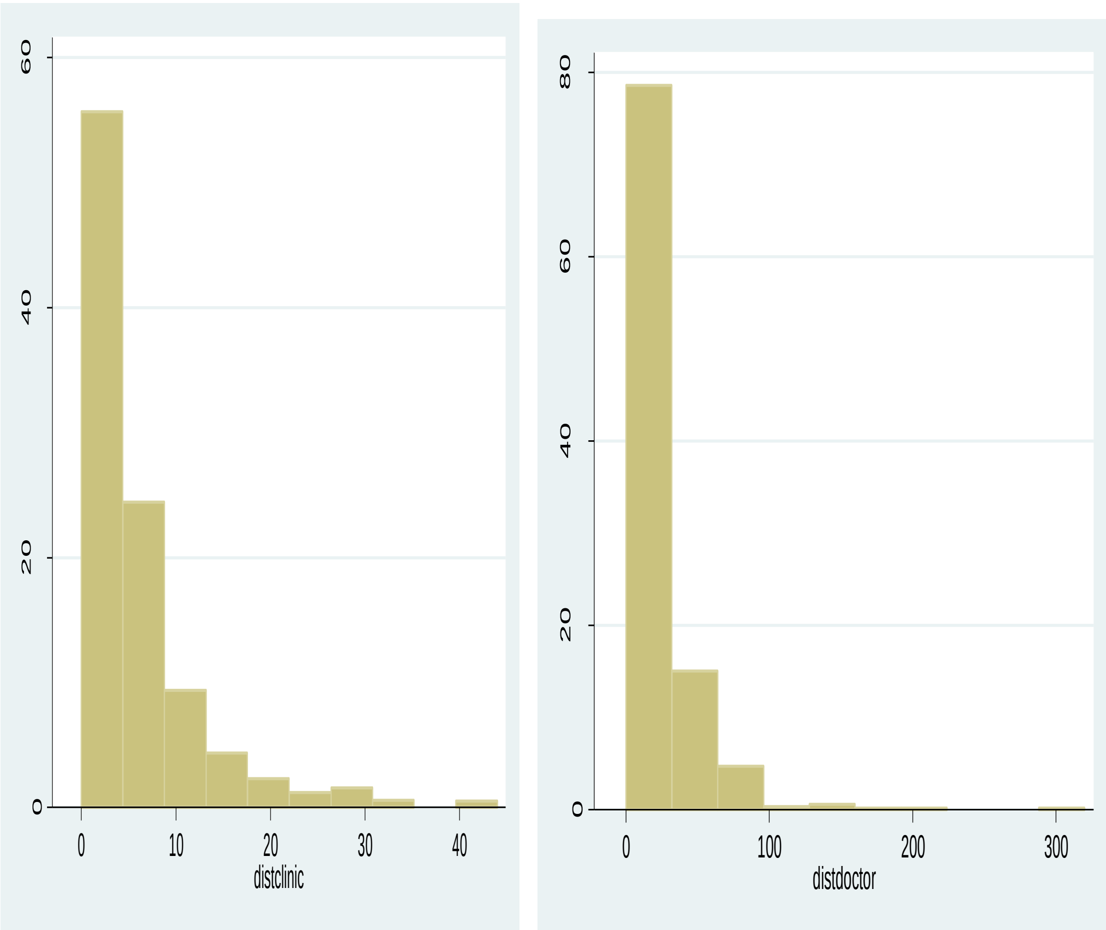
Therefore, as Lee (1982) illustrates, health and wages are jointly determined, with health appearing in the wage equation and wages also appearing within the health equation.[162] Approaches that may be considered when dealing with such simultaneous equation models are a system or full information model and/or a limited information estimator such as instrumental variable (IV) estimation. A full information model would require that estimations for both health and wages are undertaken, where wages and health (as proxied by disabled days) are both determined within the system and are therefore classed as endogenous.

One benefit of the systems approach is that it is more efficient than a limited information model by the very fact that it uses more information. The challenge with this approach to simultaneous equations is that more information is required and this information may not be available within the dataset to enable each equation to be identified. Shultz and Tansel (1997) consider indicators such as the years of schooling completed, experience, the value of household assets, local food prices for key purchases, health infrastructure, community health problems and parental background characteristics within their equations. The IHS2 has some information on such indicators but not to the same degree; community health problems are not identified for example and the identification of individuals' parents is also problematic as previously noted as individuals' relationships are connected to the household head, rather than to one another. Besides the challenges of fully specifying both equations, a system estimation has the problem that if one equation is misspecified, the estimates for both equations will be biased. For these reasons, a limited information estimator may also be considered.

[162] In addition variables such as; the level of education achieved, age and location are also included.

In contrast to a systems or full information model, a limited information estimator uses just one equation. In order to deal with the endogeneity of certain variables such as disabled days IV estimation may be utilised. The advantage of this approach is that it does not require as much knowledge about the variable of concern (disabled days) as would be required in a system estimation. What is important in an IV estimation is that appropriate instruments for disabled days are selected: variables that are correlated with the measure of health which is being instrumented but not correlated with the error term. In a system approach, in principle stronger assumptions about the determinants of disabled days would also need to be made. This would include selecting the appropriate functional form for disabled days and selecting appropriate variables which would explain disabled days.

Figure 5.2.4 Distance to a clinic and doctor for all types of paid working individuals



Source: IHS2

Table 5.2.1 Summary list of variables

Variable	Variable definition
Disabled	dummy: equal to 1 if individual experienced at least one disabled day
Ddays	total number of disabled days experienced
Chronic disabled	dummy: equal to 1 if experienced at least one disabled day due to a chronic illness
Non-chronic disabled	dummy: equal to 1 if experienced at least one disabled day due to a non chronic illness
Chronic ddays	total number of disabled days experienced due to a chronic illness
Non-chronic ddays	total number of disabled days experienced due to a non chronic illness
Primary	dummy: equal to 1 if individual has primary level education
Secondary	dummy: equal to 1 if individual has secondary level education
Tertiary	dummy: equal to 1 if individual has tertiary level education
Educdummy	dummy: equal to 1 if education was missing
Experience	Post schooling experience in years
Experience^2	Post schooling experience in years squared
Expdummy	dummy: equal to 1 if experience was missing[163]
Rural	dummy: equal to 1 if resident in the rural area
North	dummy: equal to 1 resident in the northern region
Centre	dummy: equal to 1 if resident in the central region
Private	dummy: equal to 1 if working for a private firm
Government	dummy: equal to 1 if working for the government
State	dummy: equal to 1 if working for a state owned enterprise
Public	dummy: equal to 1 if working for MASAF or other public works
Other	dummy: equal to 1 if working some other type of firm
Church	dummy: equal to 1 if working for a church based organisation
Political	dummy: equal to 1 if working for a political organisation
Distance clinic	distance to the nearest health facility in km
Distance doctor	distance to the nearest doctor in km

Source:IHS2

Table 5.2.2 and Figure 5.2.4 show the potential instruments for disabled days which are initially included within the estimation: the distance to the nearest clinic and to the nearest doctor.[164] In 90 percent of cases, individuals who have had some form of employment within the last year have a clinic within 14 kilometres of where they live and within 50 percent of cases the clinic is within 3 kilometres of their home. The distance one has to travel to a doctor is slightly higher in the majority of cases with 50 percent of individuals travelling just over 12 kilometres, but a quarter have to travel over 28 kilometres and a very small proportion of individuals need to travel over 100 kilometres.

[163] In a similar approach to that used for education.
[164] Additional instruments are subsequently explored.

Table 5.2.2 Descriptive statistics of the variable given some form of employment in the 12 months prior to the survey (N= 12,247)

Variable	Mean	Median	Min	Max	Standard deviation	N	Missing
Age[165]	30.820	28	5	97	14.618	12,245	1
Wage rate	220.173	92.4	0.00	24987	805.197	3846	114[166]
Log wage	4.587	4.526	-0.555	10.126	1.094	3845	114
Ganyu rate	108.113	50	1	40000	630.052	9257	0[167]
Log ganyu	4.043	3.912	0	10.597	0.872	9257	0
Disabled	0.183	0	0	1	0.387	10559	1688[168]
Ddays	0.826	0	0	14	2.206	10559	1688
Chronic disabled	0.044	0	0	1	0.205	10557	1690
Non-chronic disabled	0.139	0	0	1	0.346	10557	1690
Chronic ddays	0.217	0	0	14	1.230	10557	1690
Non-chronic dddays	0.608	0	0	14	1.899	10557	1690
No education	0.723	0	1	1	0.448	9367	2880
Primary	0.107	0	1	1	0.309	9367	2880
Secondary	0.151	0	0	1	0.358	9367	2880
Tertiary	0.020	0	0	1	0.138	9367	2880
Experience	15.560	12	0	79	12.516	7442	4805
Experience^2	3.987[169]	1.44	0	62	6.067	7442	4805
Rural	0.880	1	0	1	0.325	12247	0
North	0.106	0	0	1	0.308	12247	0

Source: IHS2

A system approach may have potential efficiency gains over IV estimation provided the additional structure for the additional equations can be estimated. However, there are robustness advantages to IV as IV does not require strong assumptions about the determinants of disabled days to be made and the equation for disabled days does not have be estimated. Furthermore, it is useful to first examine the IV results to assess

[165] People aged 5 years and over were asked about their activities. A very small number of children (2.37%) less than 15 years of age indicated that they had worked for a wage within the last year and are therefore included within the sample above. A larger proportion of children less than 15 years of age indicated that they had been engaged in ganyu labour (12.55%)

[166] Only individuals that indicated that they were employed within the last 12 months could report a wage given the questions used. Missing data on the last payment or the period of time that the last payment covered prevented a daily wage rate from being calculated in a small number of cases. A total of 3,961 individuals indicated that they had been formally employed within the last 12 months.

[167] Everyone who indicated that they had been engaged in ganyu labour provided an indication of their daily ganyu rate.

[168] Disability status is missing in 510 cases when considering formal waged employment (including those who do not actually have a reported daily wage rate). 1,330 cases are missing when considering ganyu labour. Please also note that individuals may be employed formally and as ganyu labour, hence why the values stated exceed the missing number of cases reported in the table). .

[169] Experience squared has been scaled by 100.

whether the instruments selected for disabled days are appropriate. If disabled days cannot be suitably instrumented then there are no gains to following a systems approach, as this would require further information which would not be accessible to ensure that the determinants of disabled days are appropriately identified.

Table 5.2.2 Descriptive statistics of the variable given some form of employment in the 12 months prior to the survey (N= 12,247) continued

Variable	Mean	Median	Min	Max	Standard deviation	N	Missing
Centre	0.411	0	0	1	0.492	12247	0
Private individual	0.371	0	0	1	0.483	3827	134[170]
Private	0.307	0	0	1	0.461	3827	134
Government	0.144	0	0	1	0.351	3827	134
State	0.042	0	0	1	0.201	3827	134
Public	0.081	0	0	1	0.273	3827	134
Other	0.006	0	0	1	0.077	3827	134
Church	0.028	0	0	1	0.166	3827	134
Political	0.020	0	0	1	0.140	3827	134
Distance clinic	5.356	3	0	44	6.923	12190	57
Distance doctor	21.098	12	0	320	28.096	12120	127

Source: IHS2

5.2.2 Additional considerations

What is clear from both the previous discussion and *Chapter 2* is that there are some problems associated with investigating health and the labour market. A problem commonly highlighted within the health and productivity literature is the issue of identifying causality (Thomas and Strauss, 1997). As a result of this, Thomas and Strauss (1997) believe that few studies are conducted which examine the health returns to wages in contrast to other studies such as the returns to education.

One issue which is considered is whether it can be inferred that morbidity has an effect on wages given how wages are reported. Individuals report their wage by indicating the

[170] The type of employer is only reported if the individual is formally employed which is true in 3,961 cases. The reported number of missing cases is therefore only applicable to this sub-set of the sample considered within the table.

amount of their last payment and the period of time that this payment covers. It is then possible to calculate an average daily wage rate for the individual using this information. What is not known is when the "last" payment was made in relation to any recently reported morbidity. The last payment could have been very recent in which case the incidence of morbidity and wage payments could coincide/overlap in some way. In such a case then the causality between recent morbidity and wages should be quite clear. However, causality is potentially not so clear if in fact the individual reports a wage that was received several months ago. A recent bout of morbidity should not have any effect on a past wage if this current morbidity was a unique incident.

Shultz and Tansel (1997) reason that using a duration of illness (in their case four weeks), indicator rather than solely the occurrence of illness is a better measure of morbidity. If someone is generally in poor health then there is a greater probability that they may experience a period of morbidity in a two week period than someone who is generally in good health (Murrugarra and Valdivia, 1999). Therefore, one could therefore think of a disabled day occurring within this two week period as an indicator of health status more generally. Hence a direct overlap between disabled days and payments is not required if one makes the assumption that a current bout of morbidity is indicative of general health status.

5.2.3 Incidence of illness in Malawi

The 2005 report by the Malawian National Statistics Office (NSO) indicates that 26 percent of the sample from the IHS2 reported an illness or injury. Of those individuals reporting periods of morbidity; fever and malaria was the most common complaint followed by respiratory related illnesses. The results from the NSO do not indicate any patterns of illness across per capita expenditure quintiles, although there is evidence of variation in incidence of illness across the three regions. The highest incidence of fever/malaria related illnesses are reported in the central region whilst the highest rates of upper and lower respiratory problems (which includes tuberculosis) are reported in the northern region. The southern region reports the highest incidence of asthma and other diseases (NSO, 2005).

Table 5.2.3 Average number of disabled days due to illness or injury in the last 2 weeks, by sex, age, education and rural-urban residence: Malawi 2004 (number of survey respondents in parentheses)[171]

Sex, age and rural/urban residence		Educational level				
		All levels of education[172]	None	Primary	Secondary	Tertiary
All males						
Age 15-39	Rural	0.73 (3763)	0.79 (2232)	0.48 (409)	0.49 (514)	0.69 (16)
	Urban	0.45 (695)	0.71 (206)	0.56 (96)	0.26 (310)	0.38 (66)
Age 40 or more	Rural	0.90 (1337)	0.95 (671)	1.04 (294)	0.55 (103)	0.00 (21)
	Urban	0.28 (236)	0.36 (78)	0.17 (46)	0.03 (64)	0.08 (24)
All females						
Age 15-39	Rural	1.01 (2255)	1.02 (1274)	0.67 (117)	0.79 (135)	0.33 (6)
	Urban	0.80 (270)	0.70 (82)	1.96 (25)	0.63 (110)	0.66 (29)
Age 40 or more	Rural	1.52 (857)	1.44 (264)	1.09 (11)	0.13 (16)	0.50 (6)
	Urban	0.67 (73)	0.32 (19)	0.00 (6)	0.09 (23)	0.00 (5)
Males in wage labour force						
Age 15-39	Rural	0.71 (1286)	0.79 (640)	0.45 (124)	0.53 (298)	0.69 (16)
	Urban	0.40 (513)	0.60 (123)	0.51 (70)	0.24 (248)	0.38 (61)
Age 40 or more	Rural	0.84 (632)	0.88 (286)	1.07 (83)	0.40 (89)	0.00 (21)
	Urban	0.20 (190)	0.31 (62)	0.20 (40)	0.04 (55)	0.10 (21)
Females in wage labour force						
Age 15-39	Rural	1.04 (390)	1.13 (171)	0.74 (23)	0.83 (95)	0.33 (6)
	Urban	0.74 (189)	0.39 (38)	2.07 (15)	0.67 (100)	0.66 (29)
Age 40 or more	Rural	1.40 (121)	2.74 (34)	0.00 (2)	0.13 (16)	0.50 (6)
	Urban	0.33 (45)	0.00 (4)	0.00 (6)	0.09 (23)	0.00 (5)

Source: IHS2

Table 5.2.3 presents a breakdown of the number of the average number of disabled days reported by respondents by age, gender and residence. Of the 3,961 individuals which indicated that they had been formally employed in the 12 months prior to the survey,[173] 573 individuals indicated that they had experienced at least one disabled day in the two-week reference period whilst 2,878 indicated that they had experienced no morbidity in the same period.[174] A slightly higher proportion of ganyu labourers (who could also be formally employed) experienced disabled days. A total of 1,578 individuals experienced at least one disabled day which represents almost a fifth of all ganyu workers.[175]

[171] Does not include cases where disability status was missing.
[172] Not inclusive of cases where education is missing.
[173] 3,961 individuals includes all those who indicated that they were formally employed even if a daily wage rate was subsequently not calculated due to missing information.
[174] Disability status is missing in 510 cases.
[175] Where disability status was reported. For reference, disability status was missing in 1,330 cases for ganyu labourers.

The average number of disabled days results cannot be readily compared to those reported by Shultz and Tansel (1997) for Ghana and Côte d'Ivoire as the reference periods are different; two weeks in the case of Malawi and four weeks for Shultz and Tansel (1997). However, one can consider the overall incidence of illness to see if there are some similarities in terms of the age, gender and education profile between the countries. Taking the top portion of Table 5.2.3 it is observed that on average, all rural males across all levels of education[176] experience more disabled days than their urban counterparts, which is similar to the patterns observed by Shultz and Tansel (1997).

Assuming that those residing in the rural area are further away from medical facilities, one may expect that those in the rural area will experience longer periods of morbidity as accessing medical assistance would be more difficult and potentially more costly due to transportation issues. Moreover, it is also observed that generally the average number of disabled days for both men and women is higher amongst those aged over 40, in both rural and urban areas across all levels of education. Shultz and Tansel (1997) found that in both Ghana and Côte d'Ivoire, those aged 40 or over experience more disabled days than those aged less than 40 with the same characteristics.[177] The finding that those aged over 40 experience more disabled days than their younger equivalent is not unexpected. As we age, health naturally declines and so we would expect that older people would be more likely to experience periods of illness. Given that the life expectancy at birth of a Malawian is 53 years,[178] 40 years old is relatively old for a Malawian (World Bank, 2012).

The mean number of disabled days reported by men and women are found to be unequal[179] using a t-test. This result is found using a simple comparison but is also confirmed when using additional controls such as the area of residence and age. Specifically considering waged workers, there is also no evidence that the mean number of disabled days reported by men and women are the same, as indicated by large t-statistics using a range of different comparisons.

[176] With the exception of rural males with primary level education aged 15-39 years old.
[177] Referring to the entire sample.
[178] 2009 estimate.
[179] Women experience more disabled days. Specifically considering male wage workers, the average number of disabled days experienced is 4.24. Female workers on average experience 4.49 disabled days.

Chronic illness is self-reported within IHS2. Respondents are asked *"do you suffer from a chronic illness?"* which is followed by a question which asks them to identify their condition and how long they have suffered from it. It is noted in the field supervisors manual that chronic conditions are specifically referred to *"medical conditions that individuals have over a long period of time (such as more than 6 months), such as diabetes. Epidemics which regularly hit the community repeatedly should not be counted as chronic illness."*

Clearly, this question is potentially subject to reporting error as an individual could self-diagnose their condition. In total out of the 12,247 individuals who did work (for a wage or as ganyu labour) in the year prior to the survey, 1,492 indicated that they suffered from a chronic illness. Of these 1,492 individuals, the diagnosis of the chronic illness is only indicated in 675 cases; in almost a third of cases this is not by someone with a medical background. Most commonly, people self-diagnose (15% of cases) or have another household member diagnose their condition (9%). However, there is some indication of the types of chronic illness that people suffer from in 99% of cases where an individual indicated that they suffer from a chronic illness. These conditions are reported in Table 5.2.4.

The reliability of this variable should therefore be considered with some degree of caution. All other people experiencing morbidity within the sample are classed as non-chronically disabled. However, given that there is potential reporting error for chronically disabled individuals, by default, there is also potential error in classifying non-chronically disabled individuals as non-chronically disabled. There is insufficient information within IHS2 to measure such an indicator with greater accuracy. Therefore, such concerns should be noted when reviewing the results pertaining to these indicators.

As Table 5.2.4 shows the most commonly reported chronic health condition for workers (both waged and ganyu labour) is arthritis, affecting a fifth of all those experiencing chronic health problems followed by asthma and more severe malaria related health problems.

Table 5.2.4 Types of chronic illnesses reported by working (formal and ganyu) respondents

Type of chronic illness:	Frequency	Percentage
Chronic malaria/fever	205	13.90
TB	40	2.71
HIV/AIDS	5	0.34
Diabetes	13	0.88
Asthma	174	11.80
Bilharzia/Schistosomiasis	86	5.83
Arthritis/rheumatism	325	22.03
Nerve disorder	43	2.92
Stomach disorder	151	10.24
Sores that don't heal	58	3.93
Cancer	3	0.20
Pneumonia	66	4.47
Other	72	4.88
Pain of limbs, joints or swellings	53	3.59
Skin problem	4	0.27
Fainting/epilepsy	33	2.24
Deafness	1	0.07
Dental problem	16	1.08
Problem with internal organs	11	0.75
Sexually transmitted infection (STI)	12	0.81
Heart related problems/High blood pressure	85	5.76
Mental illness	1	0.07
Eye sight problems	18	1.22
N	1,475	*100*

Source: IHS2

A tiny proportion of those experiencing chronic health problems attribute them to HIV/AIDS. Given that the HIV prevalence rate in Malawi is 11% for those aged 15-49[180] (World Bank, 2012) one might expect that the number of people experiencing HIV/AIDS would be higher. However, the small number of individuals reporting HIV/AIDS is not particularly surprising given that HIV status may often be unknown due to a lack of formal diagnosis and the social stigma (Barnett and Whiteside, 2006) associated with HIV/AIDS may preclude people from reporting this particular health problem even in cases where is it confirmed.

Of those who have worked in both waged or ganyu activities within the year prior to the survey, just over four percent indicated that they had experienced at least one disabled day as a result of suffering from a chronic illness. The average number of disabled days

[180] 2009 estimate.

for waged workers experiencing chronic illness is 4.86 days, whereas for someone suffering from a non-chronic illness the average number of disabled days is 4.13.[181] If only ganyu workers are considered, the mean number of disabled days associated with chronic illness is 4.80 and 4.40 for non-chronically disabled individuals.[182]

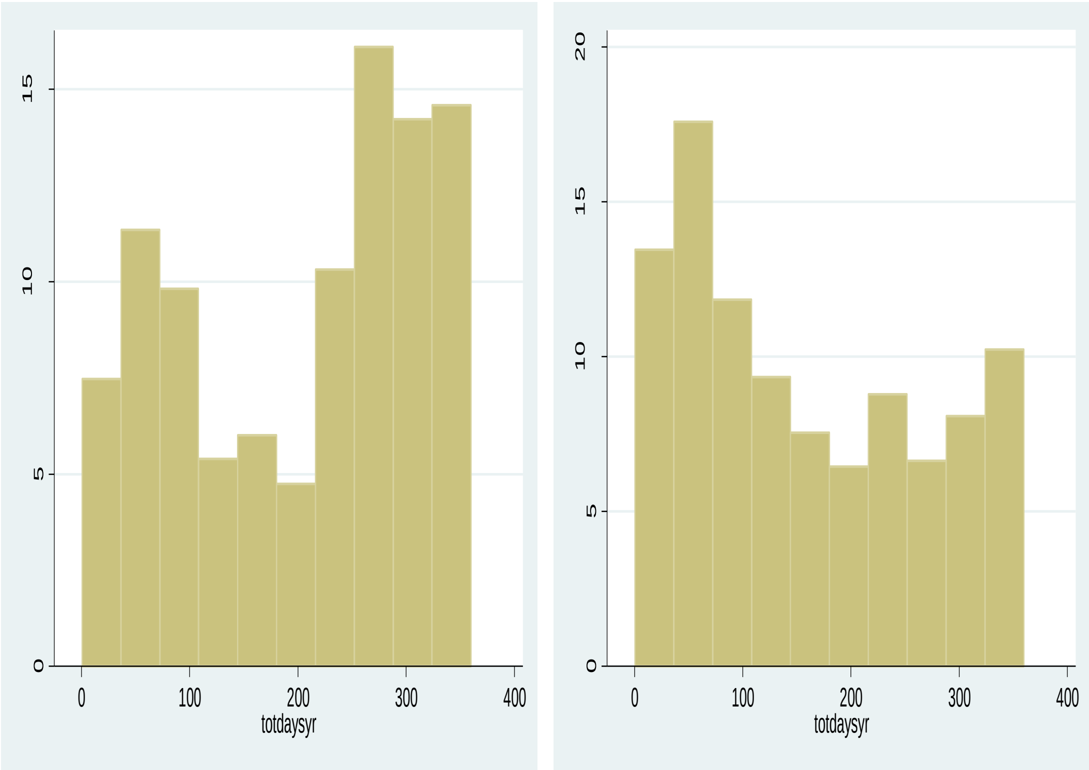
5.3 Regression estimate results

Given the discussion in section 5.2, in the first instance a limited information model will be used to estimate the impact of morbidity which is representative of a reduction in an individual's stock of human capital upon wages. Morbidity will be explored in a number of ways within this chapter, including the incidence of disability, the number of disabled days and the incidence of chronic and non-chronic disability. Individuals were asked to report all their sources of income; wages, in kind payments and ganyu (casual labour) income. Alternative measures of daily income will be considered to see if the impact of morbidity depends on the form of income the individual receives. The dependent variable used within the models is the daily wage rate[183] expressed in natural logarithms. Only those individuals who receive some form of wage from employment have been selected as I believe that waged workers are more likely to report morbidity accurately as suggested by Shultz and Tansel (1997). In total 3,846 individuals reported a wage rate and the disability status of these individuals is reported in 3,351 of cases.

It is hypothesised that the average daily wage will be lower for those who have experienced a disabled day as those experiencing morbidity will possess a smaller stock of human capital ceteris paribus relative to those in good health. As a result of this, those who experience morbidity may be less able to gain a well-paid job and/or maintain a well-paid job. Figure 5.3.1 shows the approximate[184] number of days per year that waged individuals worked; it is clear that individuals who have not experienced any morbidity (left hand panel) tend to work more days per year relative to those who have experienced morbidity (right hand panel).

[181] The median number of disabled days is 4 and 3 respectively.
[182] The median number of disabled days is identical for ganyu workers is 3 for non-chronic individuals and 4 for chronic individuals. both waged and ganyu workers.
[183] This can be formal wage income, in-kind income, ganyu income or some combination of the three sources of income.
[184] Calculated using the number of days worked in the past month and the amount of time the individual was employed for within the last 12 months.

Figure 5.3.1 Distribution of the number of days worked in a year for healthy and "disabled" individuals in waged employment

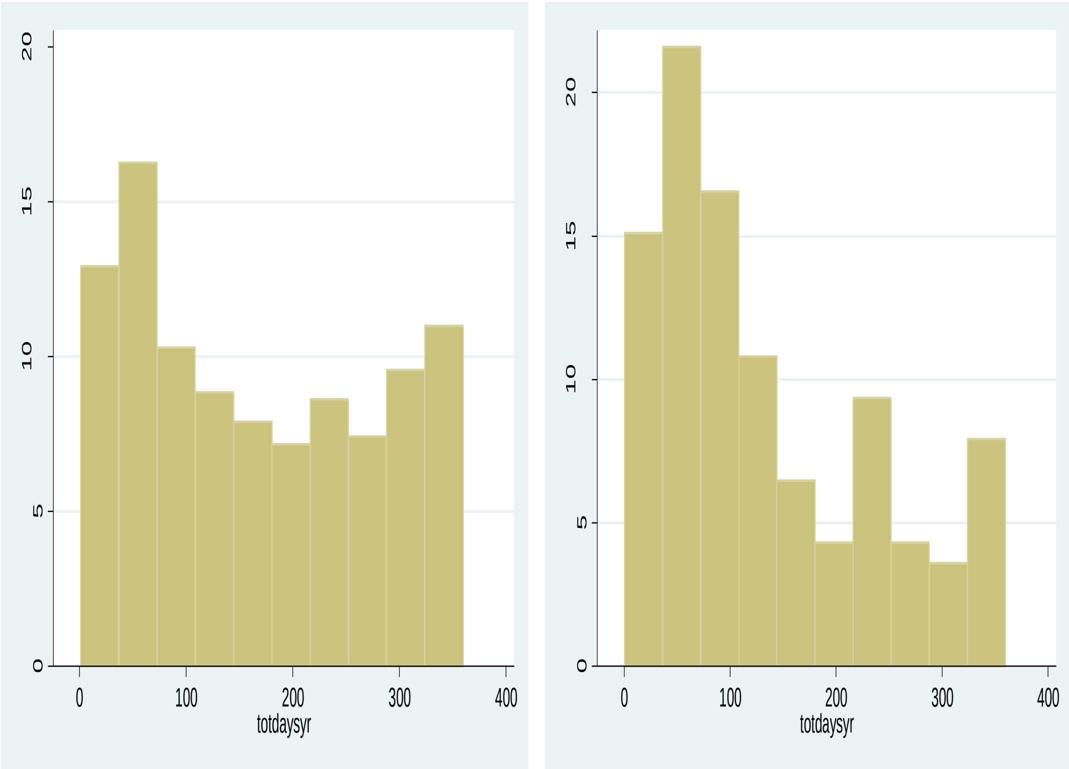


Source: IHS2

It is hypothesised that additional disabled days further erode an individual's human capital stock, resulting in a lower daily wage rate. Moreover, the hypothesised impact of morbidity upon wages will be greater when the period of morbidity is associated with a chronic illness.

Chronic illnesses may be degenerative over time and further deplete an individual's stock of human capital. Figure 5.3.2 shows the approximate number of days worked per year by individuals experiencing chronic and non-chronic health conditions. In comparing the two panels of Figure 5.3.2, it can be observed that a larger proportion of non-chronically disabled (left hand panel) individuals work almost a full year relative to those who are chronically disabled (right hand panel). A much larger proportion of chronically disabled individuals work for under a third of a year compared to non-chronically disabled individuals. A comparison of the left hand panel of Figure 5.3.1 and the right hand panel of Figure 5.3.2 is even more striking: people who are chronically disabled work far fewer days per year than healthy individuals.

Figure 5.3.2 Distribution of the number of days worked in a year for non-chronic and chronically disabled individuals in waged employment



Source: IHS2

5.3.1 The impact of morbidity on the daily wage

The impact of experiencing at least one disabled day in the two week reference period on the daily wage rate is first considered. A Chow test can be used to indicate whether there are any differences between men and women and whether, therefore, the two groups could be pooled together. The results from this test indicate a rejection of the null hypothesis, therefore pooling them is not appropriate and they will be examined separately. A similar result is found when considering the number of disabled days in Table 5.3.2 and therefore the results for men and women are reported separately.

The results in Table 5.3.1 below show that the daily wage is affected by an individual experiencing at least one disabled day. All individuals aged over 5 who reported a wage have been considered rather than specifically prime aged individuals (15-49 years of age) in order to maximise the number of observations available. If the sample were restricted to only prime-aged individuals, 478 observations for males and 135 observations for females would be lost. There are some small differences in the magnitudes of the coefficients if only prime-aged individuals are considered as opposed to all those working in the waged employment sector. However, the statistical significance of each of the coefficients does not change. The t-statistics are very large in general whether prime-aged individuals or all individuals are used. Therefore to maximise the sample size, all individuals are used within the estimations and the discussion that follows.

Looking at Table 5.3.1 column (2) it is observed that a male experiencing at least one disabled day in the two week reference period will receive on average a daily wage which that is eight percent lower than their healthy equivalent. This is a large effect and it is also statistically significantly different from zero at the 1 percent level. The 95% confidence interval indicates that the effect of being disabled is large: indicating a reduction in the daily wage rate of between 17.1 percent and 0.3 percent. For women, the effect is even less precisely estimated and not statistically significantly different from zero in column (4).

The coefficients on the additional independent variables are as expected. The rate of return to all levels of education relative to the omitted benchmark of "no education" is very large and fairly precisely determined for both males and females. Using column (2) from Table 5.3.1, it is observed that possessing primary education increases the daily wage received by 33 percent relative to someone with no education. The 95 percent confidence interval for primary education in column (2) of Table 5.3.1 suggests that at a minimum the gain to primary education is a daily wage that is at least 21 percent higher than someone with no education. This is a considerable effect, especially in a country where relatively few people possess education. The results indicate that there are considerable gains to be made from investing in education. As expected, higher levels of education are associated with higher wage rates. It is also interesting to note that males with tertiary level education also work on average 254 days per year in comparison to their counterparts with no education, who work 193 days on average per year. A closer examination of the characteristics of the two groups also shows strong differences in location; a substantial proportion of tertiary level males live in an urban area, 68 percent, compared to only 17 percent of males with no education at all.

For women, a similar result is observed in column (4) of Table 5.3.1, although the estimated effect of primary education on the daily wage is much more imprecisely estimated. To compare the estimated rate of return to possessing differing levels of education with existing literature, column (6) of Table 5.3.1 is included below. This shows that the estimated rate of return to primary level education is in excess of 35 percent relative to someone with no formal education. Psacharopoulos and Patrinos (2004) estimate that the private return to education within the sub-Saharan Africa region as a whole is around 37.6 percent and 15.7 percent for a study that used much older data from 1982 in Malawi. Moreover, the estimated return to achieving secondary education from the 1982 data is estimated to be 16.8 percent relative to someone with primary education. A very high rate of return is also observed for those possessing tertiary education: a rate of return of 46.6 percent when compared to those having achieved secondary education (Psacharopoulos and Patrinos, 2004).

The results presented in Table 5.3.1 illustrate a similar pattern to that found by Psacharopoulos and Patrinos (2004): large rates of return to primary education followed by increased rates of return as the level of education rises with the greatest returns being realised by those possessing tertiary education. Table 5.3.1 shows that the rate of return to tertiary level education for a male of average characteristics is very large: over 200 percent. Thomas and Strauss (1997) also find very large returns to education in their study of Brazil, with a 50% increase in wages being associated with someone being literate relative to an illiterate individual and rising to up to 550% in the case of someone possessing a secondary school education.

A point worth noting is that whilst there are very large returns to tertiary level education, the reality is that very few people within Malawi possess such a level of education. The IHS2 data shows that less than 5% of all waged workers within the sample have such advanced education. Amongst waged workers, secondary level education is the most commonly held level of education (putting aside those with no education), and this group represent 31% of all waged workers. The rate of return to secondary education is also very large. An individual of average characteristics with secondary education is estimated to receive ceteris paribus at least double[185] the wage received by someone with no formal education. If years of schooling are used instead of the level of schooling which is more in line with the traditional Mincer (1974) approach, the average rate of return to an additional year of schooling is estimated to be 9.6 percent for males and 13.4 percent for women[186]. This result is again similar to the results found by Psacharopoulos and Patrinos (2004) for sub-Saharan Africa as a whole.[187] They report the rate of return to an additional year of education as 11.7 percent and the mean number of years of schooling as 7.3; in the IHS2 sample for working individuals the average number of years of schooling is slightly higher at 10.3 years.

An alternative is to use the level of qualification achieved by the individual rather than years of schooling as the time spent within school may not translate to the achievement of a particular qualification. Within the IHS2, data is reported as a series of qualifications as outlined in *Chapter 3.* These have been converted into broad categories for ease of interpretation but it is also possible to use the type of qualification achieved by an individual.

The coefficients on experience and experience squared in Table 5.3.1 display the signs one would expect. Men on average have more experience than women within the sample of wage earners which is expected, given that women may leave the labour market to have children. Individuals with tertiary level education tend to possess less experience on average which is not unexpected as such individuals will have spent more time in education relative to those who have joined the labour force earlier on in life. The mean return to one year of experience for a male worker is 5.8%. An additional year of experience for an average male worker with an average of 12.4 years of experience is 3.3%[188]. For female workers, one year of experience increases the wage rate by 4.8% using column (4) of Table 5.3.1. The return to an additional year of experience for an average female worker with the average of 9.3 years of experience is

[185] The exception to this is the specification in column (2), where a male of average characteristics with secondary education will receive a wage which is 90% larger than that taken home by someone who has achieved no formal education level.

[186] 12.8 percent for all wage earners. Results not presented.

[187] A specific rate of return to additional years of schooling is not reported for Malawi.

[188] The coefficient on experience squared has been rounded up in Table 5.3.1 column (2) from -0.000927.

2.9%. The marginal effect of each additional year of experience is always falling and the declines to zero after 29 years for male workers and at 24 years in the case of female workers. Realistically, after more than 20 years of working, many people in the sample would be heading toward the end of a working life.

The effect of living in the rural area is also very large for both sexes and statistically significant in all cases. Employees within the rural area receive much lower wages: approximately half the average daily wage that their urban equivalent would receive all other things being equal. Such large differences between the rural and urban areas could be associated with differences in living costs between the two areas; if the living cost in the urban area were very high then one would expect that wages would be higher to compensate for higher living costs. However, it is equally likely that people working in the urban sector are in different types of jobs which require them to be far more productive and therefore the difference in wages is a reflection of the difference in productivity.

Malawi is interesting as it is different to Ghana and Cote d'Ivoire where the majority of wage earners live in the urban area according to Shultz and Tansel (1997). Instead the IHS2 indicates that in Malawi, 73 percent of wage earners live in the rural area and only 27 percent live in the urban area. As an alternative, columns (1) and (3) of Table 5.3.1 include regions as controls for location. For male wage earners, living in the Northern or Central region of Malawi is associated with a daily wage rate that is lower than that received within the Southern region by just over 10 percent. In order to explore the issue of job type on the daily wage rate, differences in wages by firm type will be considered later on within the chapter. Respondents were able to provide a description of their job and subsequently there are many types of occupation which would not necessarily be easy to categorise into groups. Therefore the type of firm that an individual works for is considered as there are fewer and more clearly defined categories.

Table 5.3.1 only considers whether the individual experienced a disabled day or not in a very crude manner. In reality one would expect to observe differences in the impact of disability in relation to the length of the disability which prevents an individual from working. Given that additional disabled days reflect a greater reduction in the stock of human capital that each individual possesses, it is hypothesised that additional disabled days will have an increasing impact in reducing the daily wage rate received. Table 5.3.2 is similar to Table 5.3.1 with the exception that duration rather than incidence of morbidity is now considered.

Table 5.3.1 OLS estimates of the impact of experiencing at least one disabled day in a two week reference period on the daily wage rate for all working individuals by gender.

Dependent Variable: Natural Logarithm of Daily Wage						
	Males		Females		All individuals	
	(1)	(2)	(3)	(4)	(5)	(6)
Disabled	-0.134 (0.043)***	-0.087 (0.043)***	0.132 (0.071)*	-0.077 (0.069)	-0.156 (0.037)***	-0.110 (0.037)***
Primary	0.413 (0.047)***	0.326 (0.046)***	0.476 (0.179)***	0.318 (0.168)*	0.445 (0.047)***	0.352 (0.046)***
Secondary	1.016 (0.042)***	0.899 (0.040)***	1.457 (0.074)***	1.271 (0.071)***	1.106 (0.037)***	0.098 (0.035)***
Tertiary	2.511 (0.119)***	2.271 (0.111)***	2.892 (0.175)***	2.595 (0.158)***	2.583 (0.099)***	2.335 (0.092)***
Educdummy	0.369 (0.129)***	0.365 (0.128)***	0.676 (0.164)***	0.655 (0.158)***	0.449 (0.105)***	0.442 (0.104)***
Expdummy	-0.130 (0.130)	-0.054 (0.129)	-0.245 (0.173)	-0.159 (0.166)	-0.167 (0.106)	-0.089 (0.105)
Experience	0.058 (0.005)***	0.058 (0.005)***	0.049 (0.013)***	0.048 (0.012)***	0.058 (0.004)***	0.058 (0.004)***
Experience^2	-0.001 (0.0001)***	-0.001 (0.0001)***	-0.001 (0.0004)**	-0.001 (0.0003)***	-0.001 (0.0001)***	-0.001 (0.0001)***
Rural		-0.463 (0.034)***		-0.564 (0.076)***		-0.478 (0.030)***
North	-0.104 (0.048)**		-0.0459 (0.094)		-0.092 (0.043)**	
Centre	-0.107 (0.036)***		-0.057 (0.065)		-0.103 (0.032)	
Constant	3.816	4.126		3.900	3.733	4.054
F	F(10,2596) =128.00 (p-value=0.00)	F(9,2597) =154.55 (p-value=0.00)	F(10,741) =74.66 (p-value=0.00)	F(9,742) =88.22 (p-value=0.00)	F(10,3348) =195.84 (p-value=0.00)	F(9,3349) =234.20 (p-value=0.00)
R^2	0.4059	0.4352	0.5394	0.5757	0.4354	0.4650
N	2607	2607	752	752	3359	3359

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

In Table 5.3.2, it is observed that an additional disabled day reduces the daily wage by almost 2 percent in the case of male workers in column (1) when using the region of residence as the location control, and this finding is statistically significant at the 1 percent significance level. The estimated coefficient for disabled days (ddays) for women reported in columns (3) and (4) is not statistically different from zero; therefore there is no evidence to support the hypothesis that additional disabled days reduce the average daily female wage rate. As noted earlier in this chapter, the sample size for females is somewhat smaller than for males, with the result that the estimated coefficients for the female sample are rather imprecisely determined.

Table 5.3.2 OLS estimates of the impact of disabled days on the daily wage rate earned by gender and location.

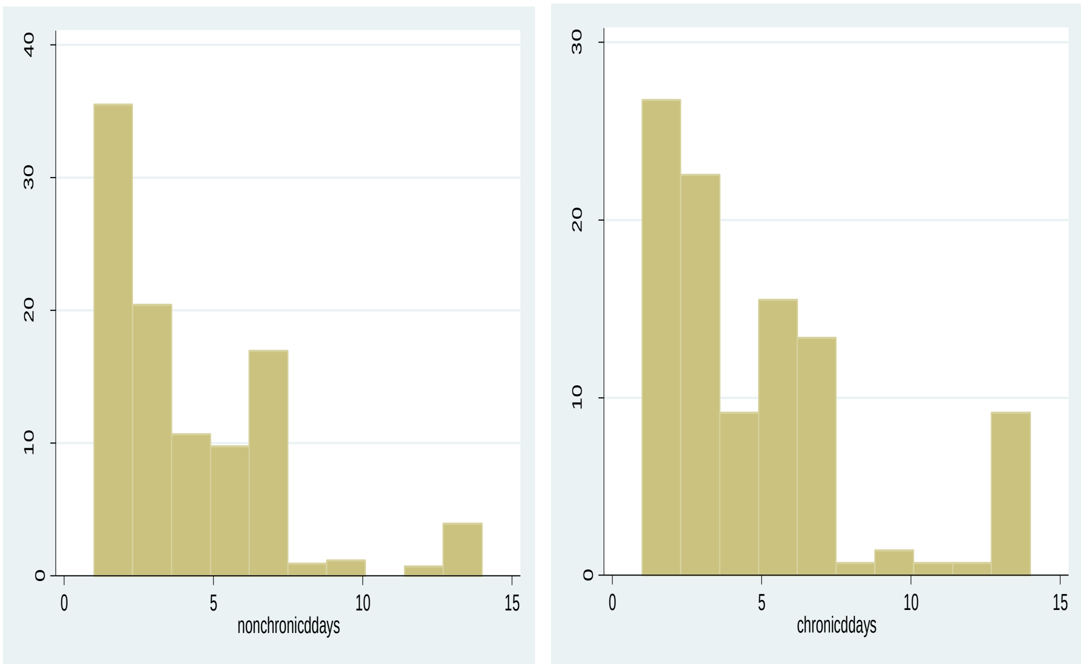
	Males		Females	
	(1)	(2)	(3)	(4)
Disabled days	-0.019 (0.007)***	-0.014 (0.007)**	-0.013 (0.016)	-0.009 (0.011)
Primary	0.420 (0.047)***	0.330 (0.046)***	0.478 (0.179)***	0.317 (0.168)*
Secondary	1.024 (0.041)***	0.903 (0.040)***	1.469 (0.074)***	1.276 (0.072)***
Tertiary	2.520 (0.119)***	2.275 (0.111)***	2.908 (0.175)***	2.600 (0.159)***
Educdummy	0.369 (0.130)***	0.365 (0.128)***	0.681 (0.162)***	0.658 (0.157)***
Expdummy	-0.128 (0.130)	-0.052 (0129)	-0.249 (0.171)	-0.161 (0.165)
Experience	0.058 (0.005)***	0.058 (0.005)***	0.050 (0.013)***	0.049 (0.012)***
Experience^{2-}	-0.001 (0.0001)***	-0.001 (0.0001)***	-0.001 (0.0003)**	-0.001 (0.0003)**
North	-0.105 (0.0477)**		-0.039 (0.094)	
Centre	-0.106 (0.036)***		-0.054 (0.065)	
Reside		-0.466 (0.034)***		-0.569 (0.067)***
Constant	3.804	4.121		3.861
F	F(10,2596) =127.33 (p-value=0.00)	F(9,2597) =154.61 (p-value=0.00).	F(10,741) =74.74 (p-value=0.00)	F(9,742) =88.00 (p-value=0.00)
R^2	0.4051	0.4350	0.5382	0.5754
N	2607	2607	752	752

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

Alternative specifications of residency as shown in columns (1) and (2) or (3) and (4) do not change the result substantially: a male of average characteristics experiencing an additional disabled day will receive a lower daily wage on average than their equivalent healthy counterpart. The largest reported effect of disabled days is 1.9% in column (1) of Table 5.3.2. For those experiencing morbidity the median number of disabled days reported in the male wage sample was four days. These individuals are estimated to receive a daily wage that is 7.6% lower than someone who had no disabled days at all.

Moreover, over a quarter of the male sample reported that they had experienced six or more disabled days. With even a relatively short period of morbidity of six days; the resulting take-home wage is estimated to be 11.6% lower than someone who had experienced no morbidity at all. At the extreme, someone who had experienced morbidity for the entire two-week reference period is estimated to receive a daily wage rate that is 27 percent less or only 73% of the comparable healthy wage. The results presented in Table 5.3.2 show that even very short periods of morbidity can have large impacts on the take-home pay of individuals. The effects and magnitude of education and experience upon the average daily wage have the expected signs in all cases and are similar to the coefficients reported in Table 5.3.1.

Figure 5.3.3 Distribution of disabled days by severity of illness for individuals who are in waged employment



Source: IHS2

The results presented in Table 5.3.2 consider disabled days without any reference to the severity of illness and therefore all disabled days are treated equally. In reality one would expect that individuals experiencing a chronic illness will experience more disabled days.

The overall distributions for chronic and non-chronic disabled days look extremely similar; the only difference is that a greater proportion of individuals who are chronically disabled experience the maximum number of disabled days and a greater proportion of non-chronically disabled individuals experience fewer disabled days, which is what one would expect.

The mean number of disabled days experienced by those experiencing a chronic illness is 4.85 days for all waged workers, whilst the mean number of disabled days experienced by those experiencing a non-chronic illness is 4.13 days[189]. The next set of estimations will follow the approach taken in Table 5.3.1 and Table 5.3.2 respectively. Table 5.3.3 presents results for males. The coefficient on chronic disabled in column (1) indicates that a male experiencing at least one disabled day as a result of a chronic illness will receive a wage that is 19 percent lower than a male who did not experience a disabled day or who is non-chronically disabled, and this is significantly different from zero at the 1% level. However, the coefficient on non-chronic disabled shows that the impact of a non-chronic disabled day on the daily wage for men is not statistically different from zero. Previously it has been argued that one potential issue with the data reported within the IHS2 is that wages and morbidity information may not coincide and are actually related to different periods of time. The wage received last month, for example, would not be affected by a period of morbidity that occurred more recently *unless* that individual generally experiences morbidity and thus has a reduced stock of health and therefore less human capital. A good example of such a case would be someone having an acute condition such as a broken leg or a headache or stomach upset (assuming a condition such as a headache or stomach upset was unrelated to an ongoing issue). In such circumstances morbidity of this kind is highly unlikely to have any impact on the wage rate received unless the two week morbidity reference period and the most recent wage overlap in some way.

In contrast, the problem of when the data for wages and morbidity refer to is less problematic for chronic health conditions *if* one takes the view that for someone with a chronic health condition, there is a reasonable likelihood that they may experience some morbidity within the reference period of two weeks, given that disabled days are a proxy for general health status *and* that they are more likely to have experienced some past morbidity. Taking this view, as a result of experiencing more frequent periods of morbidity such individuals would have a smaller stock of health capital as it is eroded by additional periods of morbidity.

[189] Male wage workers on average experience 4.20 non-chronic disabled days and 4.40 chronic disabled days. Female wage workers experience 3.98 non-chronic disabled days and 5.73 chronic disabled days on average.

Table 5.3.3 OLS estimates of the impact of experiencing at least one disabled day in a two week reference period on the daily wage rate by gender and severity

	Dependent Variable: Natural Logarithm of Daily Wage			
	Males		Females	
	(1)	(2)	(3)	(4)
Chronic disabled	-0.190 (0.073)***		-0.028 (0.113)	
Non-chronic disabled	-0.055 (0.049)		-0.097 (0.081)	
Chronic ddays		-0.021 (0.012)*		0.006 (0.011)
Non chronic ddays		-0.011 (0.008)		-0.022 (0.017)
Primary	0.329 (0.046)***	0.331 (0.046)***	0.316 (0.169)*	0.311 (0.169)*
Secondary	0.890 (0.040)***	0.904 (0.040)***	1.271 (0.071)***	1.276 (0.072)***
Tertiary	2.272 (0.0112)***	2.276 (0.112)***	2.594 (0.159)***	2.596 (0.160)***
Educdummy	0.367 (0.129)***	0.364 (0.129)***	0.656 (0.157)***	0.671 (0.150)***
Expdummy	-0.055 (0.129)	-0.053 (0.129)	-0.160 (0.165)	-0.172 (0.159)
Experience	0.058 (0.005)***	0.058 (0.005)***	0.048 (0.012)***	0.049 (0.012)***
Experience^2	-0.001 (0.0001)***	-0.001 (0.0001)***	-0.001 (0.0003)**	-0.001 (0.0003)**
Reside	-0.462 (0.034)***	-0.467 (0.034)***	-0.562 (0.068)***	-0.562 (0.067)***
Constant	4.123	4.120	3.868	
F	F(10,2596) =139.38 (p-value=0.00)	F(10,2596) =139.21 (p-value=0.00)	F(10,741) =79.30 (p-value =0.00)	F(10,741) =79.59 (p-value=0.00)
R^2	0.4356	0.4350	0.5759	0.5763
N	2607	2607	752	752

Source : IHS2 Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

The more frequent and more severe the periods of morbidity are, the more difficult it is for someone to rebuild their stock of health capital. As a result, such individuals generally have a lower stock of human capital ceteris paribus and are less able to command a higher wage in the labour market. This may explain why there appears to be no significant effect of non-chronic disability on wages.

Education and experience have the expected signs and are statistically significant in all cases for men and women; the coefficients are also broadly consistent with those observed within the specifications in Table 5.3.1 and Table 5.3.2. There is no evidence to support that having a chronic or non-chronic illness has a non-zero impact on female wages, but this is unsurprising as there is very limited evidence that morbidity affects female wages in any of the other specifications.

As observed in Table 5.3.1 column (2), the reduction in the daily wage as a result of experiencing at least one disabled day within the reference period was estimated to be in the region of 9 percent for male wage workers. Accounting for the severity of the health condition associated with the disabled day suggests that the impact of a disabled day on the daily wage is greater than initially indicated. The duration of morbidity as well as the severity is also considered following the approach from Table 5.3.2. The effect of an additional chronic disabled day is to reduce the average daily male wage received by 2.1% in comparison to the wage received by a healthy individual or someone experiencing non-chronic disabled days and this is statistically significant at the 1% level. This result is larger than the estimated impact of an additional disabled day more generally. If someone experiences morbidity generally for 4 days, the effect is to reduce their daily wage by 7.6%. By taking severity of the condition into consideration this would rise to 8.4% for the same period and at the extreme of 14 days (which is the maximum permitted number of disabled days given the IHS2 data), the effect is to reduce the average daily wage by 29.4%.

As noted following Table 5.3.1, the type of firm someone works in could have an effect on the wage rate paid. Respondents were asked to report the type of firm that they work for. Most commonly people worked for private individuals as shown in Table 5.3.4. Including additional control variables for firm type into the models presented in Tables 5.3.1-5.3.3 does not alter the estimated impact of morbidity upon daily wage rates.

For example, taking the model presented in Table 5.3.1 column (2) and adding in the additional controls for the type of firm to validate the model and using a private individual as the base category there is an improvement in the reported R-squared from 43% to 48% but there is no dramatic impact on the coefficients in terms of magnitude or statistical significance as shown in Table 5.3.5. What is interesting is the magnitude of the coefficients on the type of firm individuals are working for.

Individuals working in privately owned firms, government and state organisations are all shown to be paid a daily wage rate which is substantially higher than someone who is working for a private individual and the effects are all statistically significant at the 1% level. For example; a government employee receives a wage which ceteris paribus is

31% higher than the wage received by a comparable waged individual who is working for another private individual. An individual working in the private sector is estimated to earn substantially more than someone working for a private individual: almost 50 percent more per day. These findings are not, however, wholly surprising. For example, individuals working for another individual as a maid or as a labourer would not be in particularly well paid jobs. Ten percent of all waged workers indicate that they have been employed in farming activities in the last year and almost five percent have worked as a domestic servant.

Table 5.3.4 Type of firms that waged employees work in[190]

	All workers		Male workers		Female workers	
Type of firm	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Private company	1,173	30.65	980	33.02	193	22.06
Private individual	1,425	37.24	1,142	38.69	283	32.34
Government	551	14.40	397	13.45	154	17.60
State owned enterprise	161	4.21	140	4.74	21	2.40
MASAF or other public works	310	8.10	165	5.59	145	16.57
Church	23	0.60	17	0.58	6	0.69
Political	76	1.99	42	1.42	34	3.89

Source: IHS2

In contrast, someone working for the government or a private firm would be in a more senior job with additional skill requirements which would translate to a higher wage. An examination of the slightly reduced kinds of trades/business that people work in reveals that the government sector employs a large number of people as professionals in activities related to health, education and justice. Just over a third of waged workers indicating they work for the government are employed as teachers[191] and a significant proportion work in the medical profession.[192]

A somewhat more interesting question is whether the type of firm that an individual is working in has any bearing on the impact of morbidity on wages. Taking the specification used in Table 5.3.1 column (2) for each firm type, the results of particular

[190] The type of firm that an individual was working in is missing in 134 cases.
[191] 35.53%
[192] 6.78%

interest are presented within Table 5.3.6. The results show that experiencing at least one disabled day will reduce an average individual's wage ceteris paribus by a large amount if the individual works for a private firm or the government. The effect of morbidity on government employees is particularly large. Someone working for the government experiencing at least one disabled day (which of course could be a very short period of morbidity) receives a wage which is 27 percent lower than an equivalent healthy government employee.

Table 5.3.5 OLS estimates experiencing at least one disabled day with the addition of controls for firm type

Dependent Variable: Natural Logarithm of Daily Wage		
	Men	Women
Disabled	-0.087 (0.041)***	-0.102 (0.065)
Primary	0.292 (0.044)	0.172 (0.172)
Secondary	0.760 (0.043)***	1.044 (0.076)***
Tertiary	2.066 (0.110)***	2.213 (0.164)***
Educdummy	0.409 (0.119)***	0.519 (0.146)***
Expdummy	-0.161 (0.034)	-0.148 (0.150)
Experience	0.049 (0.005)***	0.033 (0.011)***
Experience^2	-0.001 (0.0001)***	-0.0005 (0.0003)
Private	0.480 (0.035)***	0.588 (0.076)***
Government	0.307 (0.051)***	0.380 (0.106)***
State	0.468 (0.067)***	0.816 (0.169)***
Public works	-0.226 (0.058)	0.601 (0.083)
Other	0.055 (0.207)	1.376 (0.458)***
Church	0.321 (0.100)***	0.678 (0.140)***
Political	0.564 (0.163)***	0.714 (0.134)***
Rural	-0.460 (0.034)***	-0.600 (0.068)***
Constant	4.027	3.813
F	F(16,2577)=107.04 (p-value=0.00)	F(16,731)=64.16 (p-value=0.00)
R^2	0.4817	0.6234
N	2594	748

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

Comparing the mean effect of experiencing at least one disabled day across private and government employers indicates a rejection of the hypothesis that the effect of morbidity upon the daily wage rate is the same across the two groups. The estimation results by firm type generally suggest that morbidity doesn't affect the daily wage rate for most firm types.[193] Specifically, there is no evidence to suggest that experiencing morbidity has any effect on the daily wage rate in all other firm types; private firms and government employers are the exception rather than the rule.

Table 5.3.6 OLS estimates experiencing at least one disabled day by firm type (male waged workers only)

Dependent Variable: Natural Logarithm of Daily Wage		
	Private firm	Government
Disabled	-0.135 (0.068)*	-0.270 (0.125)**
Primary	0.248 (0.075)***	0.274 (0.156)*
Secondary	0.648 (0.068)***	0.656 (0.117)***
Tertiary	2.377 (0.126)***	1.644 (0.157)***
Educdummy	0.256 (0.166)	0.310 (0.278)
Expdummy	0.078 (0.170)	0.177 (0.236)
Experience	0.038 (0.007)***	0.038 (0.012)***
Experience^2	-0.0001 (-0.001)***	-0.001 (0.0003)***
Rural	-0.341 (0.059)***	-0.323 (0.084)***
Constant	4.526	4.485
F	F(9,860)=72.55 (p-value=0.00)	F(9,344)=20.57 (p-value=0.00)
R^2	0.4316	0.3499
N	870	354

Source: IHS2. Notes: *** significant at the 1% level, **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

A large number of government employees work in areas such as education, health and justice, notably as professionals such as teachers, doctors and police officers. One might expect that such individuals would be on long term contracts that may provide sick pay and so the daily wage rate would not be affected by a period of morbidity. However, these types of individuals only account for 53% of workers who are employed by the government: there are many others in occupations which are perhaps more likely to

[193] Results for other firm types are not reported.

have contracts which may not have any protection in the face of morbidity such as those working within catering, manual labour or agriculture. An alternative or complementary explanation is endogeneity: the people in government employment who are less likely to experience morbidity are also the ones more likely to be found working as professionals such as teachers etc. I return to the endogeneity issue later in the chapter.

5.3.2 The impact of morbidity on in-kind payments and the daily wage

The vast majority of individuals in the IHS2 survey did not report any form of in-kind payment. Respondents were asked to estimate a cash value equivalent of any in-kind payment they received. Such an estimation of the value of very differing forms of in-kind payments is problematic as individuals could view the "value" of such payments very differently. In-kind payments were defined within IHS2 as additions such as uniform, housing, food and transport which were added to any wage they received. Given this definition, only those who were engaged in waged employment can receive such payments.

A total of 1,252 respondents indicated that they received some form of in-kind payment, which represents just under a third of all waged employees. The median value of in-kind payments is 50 MK. The mean is much higher at 159 MK but this is driven by a small proportion of respondents who reported very large values for in-kind payments. For the individuals receiving a wage and some form of in-kind payments, in-kind payments represent approximately two-fifths of their total daily take-home pay, which is a substantial addition to the formal wage. Given that a reasonable proportion of waged individuals also receive some form of in-kind payment and these in-kind payments are not insignificant amounts, it is worthwhile to consider the impact of illness on total daily take-home pay as well.

There is no information within the IHS2 to indicate the detail of each individual's in-kind payments. Such information would be preferable as the effect of morbidity on these additional payments is highly likely to vary by the type of in-kind payment that the individual is in receipt of. It would not be unreasonable to suggest that the relatively modest values of in-kind payments could be reflective of items such as uniform and perhaps food, whereas the larger payments are perhaps more likely to be reflective of much more generous in kind-payments such as housing.

If an individual is experiencing morbidity then items such as food and transport would not be provided to the individual (assuming food is provided as a meal in conjunction with undertaking their job) and there would be no need to transport an individual from their home to work. In such cases, the reported loss in terms of a daily rate would be larger than if for example someone was provided with housing; a short period of morbidity would not necessarily result in the loss of an individual's home immediately. However, if the individual suffered from a chronic condition, the probability of losing such a valuable in-kind payment is much more likely and would be a substantial loss to the individual concerned.

Previously it was observed that the effect of experiencing morbidity reduced the daily male wage rate by almost 9 percent and this was statistically significant at the 1% significance level. The results presented in Table 5.3.7 suggest that experiencing morbidity has no effect on the combined formal and in-kind wage whilst at the same time suggesting a turnaround in the case of women. In previous specifications, the effect of morbidity upon female daily wage rates has been statistically insignificantly different from zero; now it is observed that not only is the effect statistically significant in Table 5.3.7 but also negative and very large, indicating that women experiencing at least one disabled day will have their wage reduced by half. The confidence interval for the coefficient on disabled indicates that at a minimum, the reduction in the take home pay of women experiencing at least some morbidity will be 20 percent lower compared to a healthy female, ceteris paribus.

An examination of the types of occupation men and women reported may help to explain these results. Less than 2 percent of male workers receiving additional in-kind payments worked in occupations such as housekeeping or maid-services. In comparison, 8 percent of women worked in these occupations which may be associated with regular in-kind payments in the form of food and perhaps larger value additions such as housing which may be more commonly provided within such occupations compared to others. Moreover, a larger proportion of women experienced chronic disability within the sample of workers receiving a wage plus some in-kind benefits. Previously it was proposed that the effects of morbidity are greater in the case of chronic health conditions. This is because chronic health conditions erode health capital over time which may be difficult to replenish. Hence, such individuals consistently possess a smaller stock of health and human capital. A woman experiencing a chronic health condition who was employed as a live-in housekeeper, could find that a short period of morbidity could have a very large impact on her, if this then resulted in the loss of housing which accompanied her job. Unfortunately, in the absence of more detailed information on in-kind payments it is only possible to speculate about the very large influence of morbidity on the daily wage and in-kind rate.

Table 5.3.7 OLS estimates of the impact of experiencing at least one disabled day in a two week reference period on the daily take-home pay (formal wages plus in-kind payments) by gender

Dependent Variable: Natural Logarithm of Daily take-home pay		
	Males	Females
Disabled	-0.052 (0.072)	-0.495 (0.150)***
Primary	0.195 (0.098)**	0.798 (0.374)**
Secondary	0.982 (0.063)***	1.501 (0.159)***
Tertiary	2.255 (0.133)***	2.360 (0.191)***
Educdummy	0.469 (0.185)**	0.671 (0.304)**
Expdummy	-0.128 (0.185)	-0.163 (0.284)
Experience	0.051 (0.008)***	0.028 (0.019)
Experience^2	-0.001 (0.0002)***	-0.0001 (0.001)**
Reside	-0.533 (0.065)***	-0.346 (0.157)
Constant	3.900	3.458
F	F(9,829)=75.60 (p-value=0.00)	F(9,231)=43.56 (p-value=0.00)
R^2	51.79	0.5883
N	839	241

Source: IHS2. Notes: *** significant at 1% level **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

5.3.3 The impact of illness on the daily wage received from casual employment

A form of casual employment that takes place in Malawi is what is known as ganyu labour. Ganyu labour is defined as casual labour, where people engage in activities for short periods of time, providing unskilled labour which is predominately concentrated within agriculture (Whiteside, 2000). Respondents were asked to directly report an estimate of the daily wage they received from such employment.[194] Of the 9,257 individuals who indicated that they had worked as ganyu labour within the last 12 months, 971 had also been employed as wage labourers.

Table 5.3.8 OLS estimates of the impact of experiencing at least one disabled day in a two week reference period on the daily ganyu wage by gender

Dependent Variable: Natural Logarithm of Daily Ganyu Wage		
	Males	Females
Disabled	0.011 (0.032)	-0.013 (0.029)
Primary	0.267 (0.039)***	0.203 (0.067)
Secondary	0.500 (0.057)***	0.477 (0.162)***
Tertiary	1.306 (0.370)***	omitted[195]
Educdummy	0.639 (0.038)***	0.570 (0.038)***
Expdummy	-0.401 (0.042)***	-0.337 (0.056)***
Experience	0.034 (0.004)***	0.026 (0.005)***
Experience^2	-0.001 (0.0001)***	-0.0004 (0.0001)***
Reside	-0.606 (0.057)	-0.501 (0.099)***
Constant	4.550	4.194
F	F(9,4601)=102.79 (p-value=0.00)	F(8,3307)=41.08 (p-value=0.00)
R^2	0.2031	0.1167
N	4611	3611

Source: IHS2

Notes: *** significant at 1% level **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

It is clear for the vast majority of individuals who work as ganyu labour, that this is not a regular source of income; the mean number of ganyu days worked in the past year for both genders is only 40 days with a median number of ganyu days being only 24 in a year. Only 10 percent of all ganyu individuals indicated that they worked for more than 100 days per year in this form of employment. Comparing the patterns of ganyu employment for men and women, it is observed that men work more ganyu days: 46 at the mean compared to 33 for women.

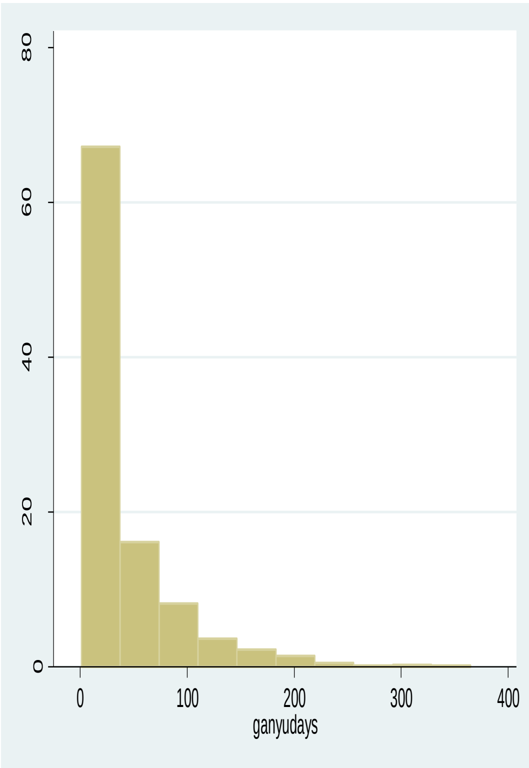
[194] Post-schooling experience is often missing. If I exclude experience and experience squared from the model used in Table 5.4.7, I would have a total of 7928 (4612 men and 3316 women respectively).

[195] There are no women with tertiary level education who also participated within the ganyu labour market.

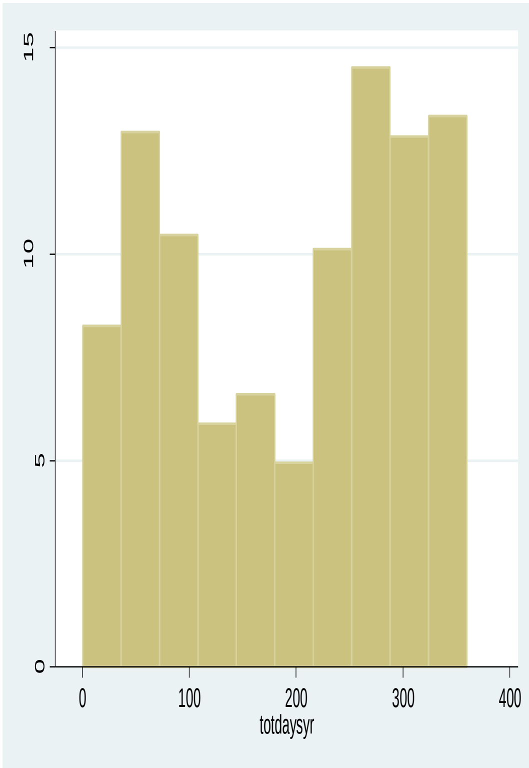
A comparison of the number of the number of days worked reveals very different patterns of employment across waged employment and ganyu employment as shown in Figure 5.3.4. There are two issues surrounding ganyu data. The first is that it may be an underreported source of income as it can be considered as shameful; despite being an important source of income for many (Whiteside, 2000).

Figure 5.3.4 Distribution of reported days engaged in ganyu labour and formal wage employment in a 12 month period

Ganyu labour days



Formal wage employment days



Source: IHS2

Moreover, individuals were asked to report *"the average daily wage you received for these days worked at ganyu over the past 12 months".* There is a possibility of reporting error as people are being asked to report information over a relatively long recall period (one year) *and* are also being asked to report an average wage for that period. This is in contrast to morbidity where the recall period is relatively short and therefore the reported information should be reasonably accurate.

People may have less recollection over not only the individual ganyu wage they received for each period of working but may also not calculate the average for this particularly accurately. The community dataset permits a rough check for accuracy of individually-reported ganyu wages: within the community dataset, community leaders were asked to report the ganyu wage for an adult male per day. They reported ganyu wages ranging between 2-500MK per day, with a mean daily ganyu rate of 92 MK and a median rate of 75MK. This is fairly close to the mean ganyu rate reported by the sample of 9,257 individuals which was 108 MK in Table 5.2.2. This would suggest that the rates reported by individuals are reasonably accurate on the whole.

Formal wages are determined by human capital factors such as the level of education an individual has achieved, their experience and their stock of health capital. Firms hire people and pay them a wage based on how productive they are; people with more human capital are more valuable to firms as people are more efficient. In principle ganyu labour is broadly similar; people would want to hire the most efficient individuals to work on their land. The difference between formal wage employment and ganyu employment is that people are temporary workers, potentially for only a few days.

An interesting point raised by Whiteside (2000) is that people may travel to different areas in order to undertake ganyu activities to avoid the social stigma that is attached to it. Therefore, an individual experiencing a period of morbidity may find that they can disguise any recent periods of morbidity by working somewhere else following their recovery; to the potential employer they appear to be as “good” as anyone else. Short periods of employment are not indicative of morbidity as the very nature of ganyu labour is that someone may only be working for a few days in any one place. Furthermore, if as Whiteside (2000) suggests workplaces are isolated from one another, provided at the time of seeking ganyu employment someone appears to be in good health, there is no reason for the potential employer to presume otherwise and the going daily rate would be paid. Therefore, it is hypothesised that the daily ganyu rate will be unaffected by short periods of morbidity.

The results in Table 5.3.8 show that the overall fit of the model is much lower than that reported in the earlier results which use the daily wage rate. In contrast to Table 5.3.1 column (2) where the included covariates explained around two-fifths of the variation in the daily wage rate of males, the same covariates only explain one fifth of the variation in the daily ganyu rate for males. Moreover, the coefficient on disabled which indicates the effect of experiencing at least one disabled day in the two week reference period is not statistically different from zero.

The results in Table 5.3.8 suggest that the factors which explain the variation in daily wage rates are considerably different to those factors which explain the variation in daily ganyu rates. Other factors such as local demand for casual labour would warrant investigation in the presence of suitable data as it is likely that given the nature of such employment that this may have a far larger influence on the daily ganyu rate. Changing the indicator of morbidity to the number of disabled days and differentiating between chronic and non-chronic disability does not alter the results in any substantial way. In all the specifications, the results indicate that morbidity health has no impact on the daily ganyu wage rate.

5.3.4 IV Estimation

As noted in section 5.1, Shultz and Tansel (1997) use IV estimation as they believe health and wages may be jointly determined. An individual who is well paid may be able to secure better nutrition, access to healthcare and so forth which in turn would enable them to increase and maintain their stock of health capital. As a result, such individuals would experience less morbidity.

If the measure of health is endogenous then the OLS estimators presented previously will likely be negatively biased: some of an apparent negative impact of morbidity on wages will in fact be driven by the negative impact that higher wages has on morbidity. Finding a variable/set of variables that is/are independent from the error term in the wage equation but correlated with the endogenous variable/s is a challenge in practice (Dufour, 2003; Angrist and Pischke, 2009). Moreover, one needs to know about the underlying relationships that determine the variables of interest: have available data on such variables and ensure that there is reasonable correlation between the chosen instrument/s and the variable/s that is/are being instrumented. Moreover, the estimated coefficients from the IV estimation can be very imprecise, and IV is not as efficient as OLS estimation.

In identifying suitable instruments, reference is made to Shultz and Tansel (1997) who use two sets of instruments, local food prices and distance to health services. They argue that the instruments in themselves will not have a direct effect on the daily wage rate but are correlated with the health variable, which in their case and in the case presented here is a measure of morbidity. The idea is that the distance an individual must travel to the local health service should have no effect on wages as they are unrelated, but the distance travelled is anticipated to be correlated with the period of morbidity experienced. Individuals with a clinic or doctor close by should in principle be able to access preventative and treatment care faster and more readily, which in turn influences the amount of morbidity that someone may experience.

In the IHS2 each individual reported the distance to their closest health clinic and doctor. If the clinic or doctor is located within the local community of the individual then the individual has no distance to travel to reach a clinic or doctor and so the distance travelled is recorded as zero in such cases. If the clinic or doctor is located outside of the local community then the distance travelled is reported in metres, kilometres or miles. The information provided by each individual has been adjusted to ensure that the distance travelled to the two health services is registered in kilometres in all cases.

One would expect that someone who is relatively close to health facilities would have better access to health services. Therefore the incidence and length of morbidity for people living close to a clinic/doctor would be lower for these individuals relative to those living further away from a clinic/doctor. It is important to note that the reported distance to a clinic/doctor does not measure the ability for someone to pay for health services, consistent with the argument that these instruments are exogenous. Moreover, the instruments used do not capture "use" of health services; this could be problematic as someone who experienced regular morbidity would use health services more often than someone who is healthy, and this of course could be affected by an individual's wage rate.

The estimation method used is two-stage least squares, i.e., instrumental variables. The first stage of the regression using the number of disabled days is presented in Table

5.3.9, and the second stage is then presented in Table 5.3.10. The first-stage results indicate that the distance an individual has to travel to a clinic is statistically significant at the 1% significance level and that this instrument is correlated with the length of morbidity which is potentially endogenous within the wage equation. However, the distance one has to travel to a doctor is not statistically significant and hence this suggests that this instrument is not correlated with morbidity.

Table 5.3.9 First stage regression: Number of disabled days with distance to clinic and doctor as instrumental variables

	Males
Primary	-0.134 (0.135)
Secondary	-0.315 (0.096)***
Tertiary	-0.290 (0.153)*
Educdummy	0.262 (0.091)
Expdummy	-0.177 (0.199)
Experience	-0.012 (0.013)
Experience^2	0.0004 (0.0003)
Reside	0.309 (0.091)***
Distance to clinic	-0.019 (0.005)***
Distance to doctor	0.0007 (0.002)
Constant	0.624
F (10, 2572)	6.748 (p-value=0.0012)
N	2583

Source: IHS2. Notes: *** significant at 1% level **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

A surprising finding in the first stage results is that the negative coefficient on distance to the clinic indicates the further away someone is from a clinic, the fewer disabled days they will experience. This is inconsistent with the motivation outlined above for the use of distance to health care as an instrument. An alternative explanation is that individuals who live far away from a clinic continue to work even when they are experiencing morbidity as travelling to obtain medical assistance could be very difficult. Put another way, individuals who live near a clinic may be more likely to visit it when ill, and this may be more likely to lead to them taking time off from work and reporting this time off as disabled days. This is somewhat speculative, however, and suggests that caution should be applied in interpreting the IV results reported.

Table 5.3.10 Second stage IV regression: Independent variable is natural logarithm of daily wage rate

	Men
Ddays	0.368 (0.165)**
Primary	0.383 (0.074)***
Secondary	1.018 (0.073)***
Tertiary	2.382 (0.142)***
Educdummy	0.273 (0.156)*
Expdummy	0.010 (0.153)
Experience	0.062 (0.007)***
Experience^2	-0.001 (0.0002)***
Reside	-0.571
Constant	3.901
N	2583

Source: IHS2. Notes: *** significant at 1% level **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

The results presented in Table 5.3.9 and Table 5.3.10 treat the measure of morbidity as endogenous. If it is actually exogenous then OLS estimation is preferable as it is more efficient than IV estimation. Testing for endogeneity[196] indicates a rejection of the null hypothesis that morbidity is exogenous in the daily wage rate model. This would suggest that the OLS estimators previously reported are not consistent and IV estimation is preferable provided that suitable instruments can be identified, which on first sight would appear to be the case.

Some additional tests are implemented to verify the suitability and the strength of the chosen instruments distance to a doctor and clinic. The first test is a test for overidentification which is used to confirm that the selected instruments are valid in the sense that the selected instruments are not correlated with the error term in the structural equation. The result from this test[197] indicates a p-value of 0.47 and thus indicates that

[196] Test of endogeneity reports a chi2 (1)= 9.180 (p=0.0024).
[197] Test of overidentifying restrictions reports a chi2(1) = 0.523 (p=0.470)

there are valid instruments for morbidity within the estimation. An additional consideration is whether the instruments selected are strongly or weakly correlated with morbidity. If this relationship is weak (the correlation is small) then the coefficient on disabled days will not be reliable and the coefficients will be biased (Wooldridge, 1999; Staiger and Stock, 1997). Therefore it is important to test for weak instruments to ensure that this problem may be avoided before inference is undertaken.

The reported first-stage F-statistic is 6.748. Staiger and Stock (1997) suggest that as a general rule of thumb in two-stage least squares estimation, a first-stage F-statistic of less than 10 is generally indicative of weak instruments. The weakness of the chosen instruments suggests the IV estimates may suffer from bias, and that caution should be applied when interpreting the coefficient estimates and also to the preceding conclusions regarding the overidentification test (exogeneity of the instruments) and the endogeneity test (exogeneity of morbidity).

There is a very large divergence between the reported OLS estimations reported previously and the IV estimates reported in Table 5.3.9. The impact of disabled days on wages is estimated to have a positive impact that is significantly different from zero, rather than a negative or nil impact as previously found and as predicted by theory. The standard errors of the coefficients are also very large, much larger than those reported using OLS estimation. For comparison the standard error on disabled days in Table 5.3.2 column (2) is 0.007, whereas in Table 5.3.9, the result is 0.165. The confidence interval for the IV estimate on disabled days is very large ranging from 0.044 up to 0.693 and very imprecisely measured.

The problems with the IV estimation suggest that alternative instruments need to be identified for estimation purposes. Within the community data there are some alternative instruments which were not initially utilised which may be considered: local prices of some foods and certain medicines and the distance someone would travel to purchase "common[198]" medicines. Community leaders were asked whether there was somewhere that common medicines could be purchased within the community; if there was somewhere then the distance travelled was recorded as zero kilometres. If

[198] The term "common" is used within the survey and the specific medicines referred to are Panadol and Fansidar (which are a painkiller and a antimalarial drug respectively).

medicines had to be purchased outside of the local community, a distance was indicated, which would be in metres, miles or kilometres. In all cases the distance travelled was converted into kilometres.

The IV results including the distance travelled to purchase medicine indicate that the instrument is not highly correlated with morbidity and therefore does not satisfy the requirement of being an appropriate instrument for morbidity. The reported F-statistic in the first stage results is again small at 1.15 and a p-value of 0.2836. The confidence interval on disabled days is very wide which further confirms the weakness of using the distance travelled to purchase medicine as an instrument for morbidity.

Local food and medicine prices were also considered. Community leaders reported prices for a series of items including; maize, rice, chicken, beef, a number of different types of vegetable and Panadol and Fansidar[199] for each community area. The prices of such foods/items could be reported in a number of different units of measurement; this included kilogrammes, piece, bunch, heap and plates.[200] An average price for each item was generated depending on the unit of measurement; for those reporting a price for maize measured in kilogrammes, the mean price was 18 MK per kilo. A mean price could also be calculated for each of the alternative measures such as pails.[201] The average price for a small pail was 281 MK within the sample. Using the sample mean price for each of the different measures, it was then possible to calculate how many pails, No 10 plates, pieces and so forth are in a kilogramme (approximately) and convert all prices into kilogrammes regardless of how they were originally reported. For the two medicines, prices were converted into prices per piece as medicines were commonly sold in pieces (which would appear to be packets rather than singular tablets).

IV estimation is then repeated including the price of some foods and medicines as potential instruments for morbidity. In a number of cases; 8% in the case of maize prices, 2 percent in the case of Panadol prices and 44 percent in the case of Fansidar prices the item may not be reported as being available within the community and

[199] These are the two specific medications for which prices can be reported.
[200] In total there were 20 different measures used.
[201] Pails could be small or large. Therefore the average price was calculated for both a large and a small pail and indeed for all different measures where appropriate.

therefore there is no price within such communities. Fansidar is an antimalarial drug which community leaders indicated was not available within the "community or nearby" at the time of the survey in just over two-fifths of cases. This would appear on first sight to contradict to an earlier question within the community survey in which only one-fifth of communities indicated that they did not have access to "purchase common medicines such as Panadol and Fansidar" which has been utilised to calculate the distance people would travel to access medicines. Given that the earlier question referred to *both* Panadol and Fansidar more generally, this would account for the difference in reported accessibility of the medications concerned; one could legitimately answer that "common" medicines *such as* Fansidar and Panadol were available locally, whilst the latter question is much more specific and specifies the two medications which may not specifically be available. Given that data was missing on the price of Fansidar in a large number of communities as the medication was not actually available it is not surprising to observe that the price of Fansidar is not correlated with the number of disabled days experienced. Moreover, as Fansidar is also an antimalarial drug, its use would not be associated with morbidity more generally and therefore this would suggest that a more general indicator of medication prices such as the one for Panadol would be a potentially more appropriate instrument.

The results reported using maize prices as an instrument for disabled days indicate that the price of maize is not correlated with the number of disabled days that individuals experience. The F-statistic from the first stage result also remains small at 1.183 with a p-value of 0.277 which is below the rule of thumb guide of 10 indicated by Staiger and Stock (1997). The 95% confidence interval for disabled days within the second stage becomes even larger to include negative values; ranging from -0.155 to 0.187 and a standard error of 0.087 on disabled days. The use of medication price (Panadol) as potential alternative instrument for morbidity also leads to a small first stage F-statistic: 0.793 with a p-value of 0.376 and in the second stage a familiar problem is again encountered.

5.4 Conclusion

This chapter has utilised a general self-reported functional measure of morbidity to estimate the effect of morbidity upon the daily wage rate received by Malawian workers. The choice of disabled days as a proxy measure for health has some limitations. The first is that due to the wording of the question within IHS2, illnesses and injuries are reported collectively. Ideally, one would only wish to consider illnesses as injuries would tend to represent more acute conditions. Such acute conditions may have a very short term effect on an individual's stock of human capital but it is unlikely unless the injury was relatively serious that the effect would be more long term. One potential way to separate injuries out from illnesses would be to use additional information about the type of illness or injury. However, not all individuals reported the ailment in question and perhaps more importantly some of the phrasing within the IHS2 makes it difficult to determine the specifics of an ailment; an eye problem could be illness or injury related for example.

In the absence of more detailed information, illnesses and injuries are considered collectively as potential sources of morbidity. Another issue for consideration is the reference period used. Each individual was only asked to report any recent morbidity; specifically in the two weeks prior to the survey. This is not ideal for a number of reasons; the first is that the reference period would not be the same for all respondents which may potentially lead to differences in reporting as malaria is more endemic during the rainy season between November to April. The survey was undertaken between March of 2004 and April 2005 and so individuals would have been interviewed during the rainy reason.

Given malaria is the second most commonly reported type of illness within the sample, the probability of experiencing recent morbidity *could* be associated with when the individuals were surveyed; leading to a potentially greater incidence and possibly duration of morbidity for individuals interviewed during the rainy season. Individuals who were surveyed outside of the rainy season would potentially be less likely to experience morbidity associated with malaria. In reality a survey of this kind could not be taken simultaneously by all respondents and each household was selected at random.

A second issue which poses a limitation to the assessment of morbidity on individuals' wages was the potential mismatch between morbidity and the wage received. The most recent wage could coincide with a recent period of morbidity but may not due to the way the questions were presented. If the individual has experienced morbidity recently (but not previously) then there should be no effect on their wage.

Previously it was argued that the payment of wages and recent morbidity do not have to coincide if recent morbidity is viewed as an indicator of general health. It can be argued that those individuals who are in good health with a “full” stock of health capital are unlikely to experience a period of morbidity within the short reference period relative to those who possess a lower stock of health capital. In such cases, individuals with a lower stock of health would be expected to receive lower wages relative to a healthy counterpart. The results from Table 5.3.3 lend support to this view that disabled days may be a proxy for general health; chronic disability does appear to have an effect on the daily wage rate, whereas there is no evidence to support the view that non-chronic health conditions have any impact on the daily wage rate.

There are two more issues which also need to be highlighted which are perhaps the most important limitations to consider. The first is that it is not possible to capture whether an individual is in fact experiencing morbidity prior to or after the reference period, in reality the total proportion of time that someone is disabled could be far greater than what is indicated within the data presented. The second issue which is related to this point is that it is not possible to identify whether morbidity has been so severe that it has led to someone withdrawing from the labour market altogether. As a result of this, the estimations may suggest a much smaller impact of morbidity than what is true in reality. If there were suitable data to identify individuals who would normally be part of the labour pool, then this latter problem could be addressed. However, as previously outlined this is not possible and therefore remains a limitation to the study.

Another problematic limitation to the study is the difficulty in identifying suitable instruments given the endogeneity of morbidity. It is clear that the initial IV estimates are unreliable given both the counterintuitive negative correlation between distance to clinic and morbidity and the imprecise coefficient estimates. As a result of this alternative instruments were considered and employed in an attempt to overcome the problem of weak instruments. One of the problems with attempting to utilise such alternatives was that these instruments were not sufficiently correlated with the endogenous variable of disabled days in the first stage of the results. Consequently, reliable IV estimates could not be obtained and a systems approach, which was originally considered due to the joint determination of wages and health, cannot be undertaken.

The OLS results presented above are similar result to the IV results found by Shultz and Tansel (1997) who find that additional disabled days result in at least a 10 percent reduction in the daily wage received in Côte d'Ivoire and 11.7% in Ghana. Moreover, using the previous IHS for Malawi (IHS1) Dorward and Mwale (2005) estimate that morbidity (specifically HIV/AIDS related) results in a wage reduction of between 5-10%.

The results presented suggest that morbidity does have a negative effect on the average daily wage men receive. However, generally there is no evidence to suggest that women's wages are adversely affected by morbidity. Rather, the negative effect of morbidity for women is evident in reduced take-home pay which includes non-wage income and payment-in-kind.

Chapter 6: Income changes and health in the KwaZulu-Natal Province

Health can be measured in a number of different ways as observed in *Chapter 2*. Consequently, quantifying the effects of morbidity is extremely challenging. Morbidity affects different people in different ways and the impact may be felt directly by the individual, but also indirectly by others within a household, the extended family, a community and society more widely. The impact of some illnesses can be mild and short-lived, whilst there are others which are severe and long-lasting, with many more in-between.

There are means of preventing and treating some ailments which can dramatically reduce the suffering of those afflicted by disease and illness. However, we also know that access to medical advancements may be limited due to factors such as geographical proximity to medical facilities or due to a lack of financial means. Consequently, large portions of society may face morbidity and its consequences.

This chapter specifically explores how health shocks impact household income given the loss of health capital following a shock, using the Grossman (1972) model as a framework for analysis.

Within the Grossman (1972) model, it is assumed that people have an initial stock of health that will depreciate over time; as one begins to age, and can be "restocked" with additional investment, the cost of which varies according to factors such as the level of education and the cost of inputs such as medical care. An individual who is in "good health" is therefore able to use their stock of health capital to produce output within healthy times. Those experiencing health shocks will experience a loss of health capital which will reduce their productive capabilities and ability to generate earnings which could be used to invest in health capital.

This chapter will utilise the KIDS dataset which, to briefly recap, is a three wave household panel dataset for the KwaZulu-Natal province of South Africa. The first survey was initially undertaken in 1993 as part of the Project for Statistics on Living Standards and Development (PSLSD) and was one of the first nationally representative household surveys for South Africa. The survey investigated issues such as poverty, socio-economic dynamics and inequality (PSLSD, 1994). Households from this original survey were then subsequently revisited in 1998 and again in 2004. This chapter uses information from the second and third waves of the survey with reference to measures of income, household characteristics and a variation of health shocks. A more detailed discussion of the dataset can be explored in *Chapter 3*.

The use of income is one way in which well-being may be examined, but that is not to say that it is the only measure; consumption or expenditure could equally be used to examine well-being. The reason for using income (both earned and unearned) in this approach is to highlight an unusual aspect of the KIDS dataset, which could not be explored if using consumption or expenditure as a measure of well-being.

The unusual feature of the KIDS dataset is that it presents self-reported income losses, which are specifically associated with shocks experienced by the household.[202] This characteristic of the dataset enables one to examine the directly reported income losses and raises some interesting questions about the distribution of health shocks upon households. The remainder of this chapter is structured as follows: *Section 6.1* presents an outline of the model used for examining the impact of health shocks on household income in the KwaZulu-Natal province in South Africa. A discussion of the variables used within the empirical estimations is also presented. *Section 6.2* presents some discussion of the data. Section 6.3 examines the impact of a health shock using regression analysis and a range of specifications for the health shock. *Section 6.4* reports results using an alternative estimation method, propensity score matching. *Section 6.5* discusses the directly reported income losses following a health shock. *Section 6.6* briefly discusses the possibility of instrumental variable (IV) estimation and *Section 6.7* concludes the chapter.

6.1 Household income and negative health shocks

This section sets out a framework to illustrate the relationship between health shocks and income. This framework is centred on the Grossman (1972) model which considers health as a form of human capital which individuals invest in as they have a demand for “good health” due to the benefits that “good health” brings.

[202] I note that shocks are reported at the household and not individual level.

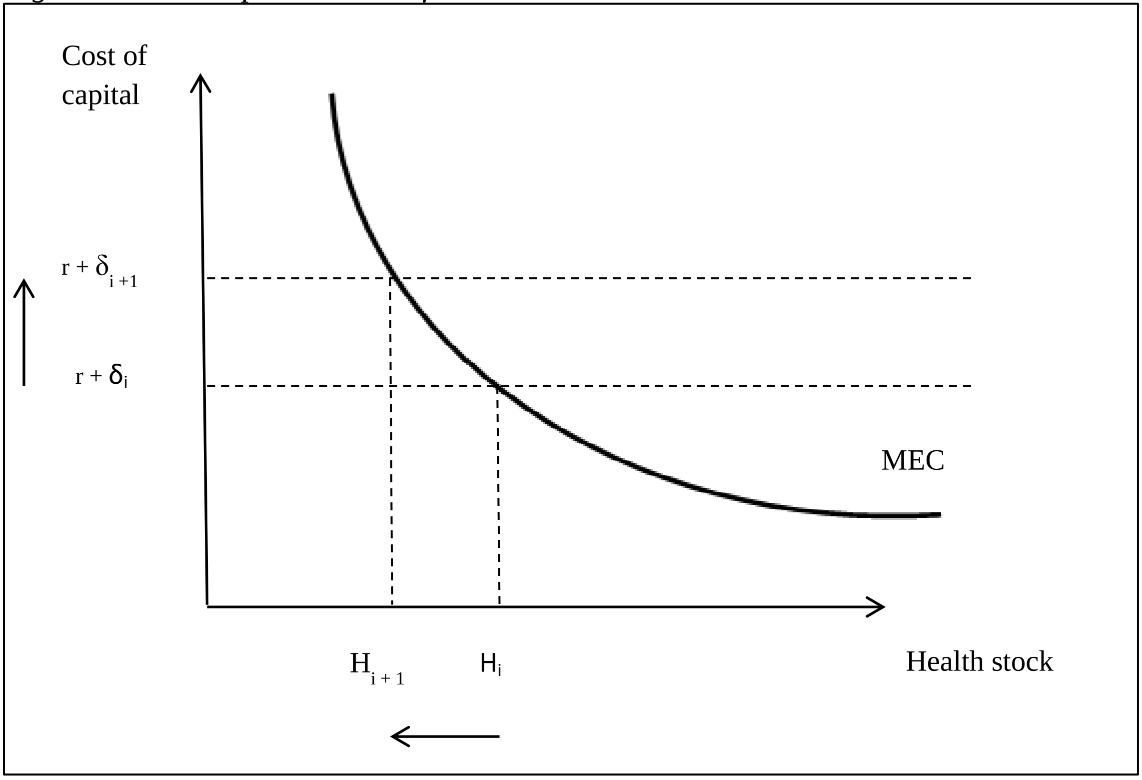
Health can be viewed as a capital good as one begins with an initial stock of health which will remain constant, grow or depreciate over time depending on the rate of depreciation within the model and the level of investment undertaken. Investment into the stock of health can be made using inputs such as a better diet, exercise, use of health care services, income and time.

People undertake such investments as there are realisable gains in the form of more favourable opportunities. This could be in the form of securing a better paid job or increased productivity off-farm which generates additional utility for the individual (Grossman, 1972).

The demand for health is determined by both consumption and investment effects (Grossman, 1972; Mushkin, 1962). It is a consumption good as it generates direct utility for individuals because the possession of good health ensures people are well and also acts as an investment; it determines how many healthy days individuals are able to undertake productive activity and therefore generate income. Over a lifetime, individuals will make choices about their optimal stock of health capital (H_i) by equating the marginal efficiency of health capital (MEC) to the cost of such investments (the rate of interest, r plus the rate of depreciation, δ) as shown in Figure 6.1.1.

The optimal choice of health stock within the model is influenced by age, wages and the education of an individual. Grossman (1972) states that as we age, the health of individuals deteriorates at an increasing rate; this is illustrated by an increase in the rate of depreciation within the model. As a result of this the optimal level of health capital demanded will fall as the effective cost of capital rises. The period of time from which one can generate returns from investments in health also falls which further reinforces the reduction in the optimal level of health stock. Higher wages have the opposite effect within the model. An individual earning a higher wage is able to experience a greater rate of return on their investment compared to a lower waged individual for healthy days. Therefore high wage individuals will demand a greater stock of health which is represented by a rightward shift of the MEC curve. The final issue considered within the model is the role of education; people with more education are predicted to be more efficient at producing health investments (Grossman, 1972). As a result of this, more educated individuals need to use fewer inputs (this could be income, exercise, diet, medical care) to produce a given level of health. The model shows that additional education raises the rate of return to an investment in health and raises the optimal stock of health (which would also be represented by a rightward shift of the MEC curve).

Figure 6.1.1 The optimal stock of health



Before proceeding it is important to draw attention to the fact that the Grossman (1972) model is based on individual health. Therefore the characteristics of each individual and their utility function can be modelled to determine an optimal stock of health which they then use for productive activity and to generate earnings. The incidence of a health shock reduces the stock of health capital that each individual possesses, the degree of which will vary from individual to individual depending on aspects such as those already noted: age, wage and so forth. With a reduced stock of health capital, the productive capacity of each individual will also decline as there are fewer healthy days available within a given time period, resulting in lower income, ceteris paribus.

This chapter utilises the KIDS dataset which reports shocks and income at the household rather than the individual level which is a significant limitation to the analysis that follows. It is not possible to identify specifically *who* within the household experienced the shock. The realisable effect of a reduction in the stock of health for someone who is nearing retirement will be much smaller when compared to someone who is in their prime earning years, for example. Unfortunately the KIDS dataset does not report any details on who experienced a health shock. Income measures are also reported at the household rather than the individual level. Therefore the approach taken within the context of this chapter is to consider household health shocks and household income.

Before proceeding with a discussion of the variables, it is important to review a couple of additional factors that need to be taken into consideration. The first issue is related to whether or not households which have been previously shocked adjust their behaviour in subsequent periods relative to what their "normal" behaviour would have been in the absence of a shock ever occurring. Households which have previously experienced a shock may seek to employ coping strategies to protect themselves against future shocks. Therefore, the estimated impact of a health shock on such households may be less than that reported by households which had never previously been shocked.

One potential way to deal with such a concern is to examine responses to previous shocks (Gan et al., 2006). An example suggested by Gan et al. (2006) is to examine previous healthcare expenditures. Therefore, we would be able to model coping strategies and subsequently obtain a better estimate of the effect of any health shock. Whilst KIDS is a panel dataset, the information collected on healthcare expenditures refers only to a short period of time. Therefore, there is not sufficient information within the datasets on healthcare expenditures to model coping strategies in the presence of shocks in such a way.[203] The approach taken by Gan et al. (2006) is to make the very strong assumption that the probability of a health shock is not dependent on previous decisions made by an individual. Therefore, no adjustment takes place in subsequent periods.

[203] There is some limited information in the 1998 KIDS dataset regarding how households coped with shocks they experienced between 1993 and 1998 using a series of dummy variables. This information, however, is not reported within the 2004 KIDS dataset.

For the 1998 data, there is some information about how the household may have dealt with shocks which occurred between 1993 and 1998. However, there is no information about the behaviour of these households in regard to previous shocks, as the shock module was only introduced in the 1998 survey. For the 2004 survey, it is known whether the household experienced a shock previously[204] but not how they responded to a shock occurring between 1998 and 2004.

A second issue is that within the estimations presented, the household head has been selected as the representative of the household, and their characteristics are reported and considered. This is similar to the approach adopted by Gan et al. (2006) who use information about the education of household head within their examination of health shocks upon income in China. There are a number of issues related to such a choice. The first issue is related to the definition of a household head. Yamano and Jayne (2004) suggest that household headship is a *"vague concept"*.[205] Within the KIDS dataset, it is clear that some of the issues identified by Rajaram (2009) such as misidentification of the head are apparent. There are a small number of cases where the respondent for the household roster identifies an individual as the resident household head, but this individual has not in fact resided in the household for any length of time during the previous 12 months. In these cases the individual coded as the absent head may be a more appropriate figure to represent the head of the household (as they have been present in the household for a length of time) rather than the one who directly identifies themselves to the survey enumerator.

An additional practical difficulty encountered when looking at the KIDS data is that in 1998 a dummy variable "head" is used to identify the head of the household along with a person code. In principle, this makes identifying the head straightforward (putting aside the previous issue of who actually is the most appropriate household head). However, in the 2004 dataset the "head" variable no longer appears, and this creates some additional difficulty in identifying the household head. In 1998, individuals who are coded with a person code of 1 are recorded as the representative resident head and someone coded with a 2 is classified as the absent head of the household respectively. In 2004, there are a number of new households which were created as a result of

[204] As households reported in the 1998 survey if they had experienced a shock between 1993 and 1998.
[205] p.94.

children from the original 1993 survey establishing households of their own. These households do not have a readily identifiable head of household as their corresponding person code is neither 1 nor 2 and the “head” variable was been dropped between 1998 and 2004. Therefore, given that the household head is not readily identifiable; the associated household has been excluded from the estimations presented. This is not ideal, but it does ensure that all households are consistently identified insofar as how each household reports their head.

An alternative representative for the household characteristics could be considered, such as the highest earning household member or the oldest working age member. An opportunity for future research is to consider whether the results are sensitive to changes in the choice of individual acting as a representative of the household characteristics. However, within the confines of this chapter, the household head has been selected.

In addition, the gender of the household head is also included as a dummy variable within the estimations as it is shown that women on average earn less than men and female-headed households are more vulnerable than their male-headed counterparts (Rajaram, 2009).

6.1.1 Income

In order to evaluate the impact of a health shock on households, one first needs to consider household income. The household acts as a natural mutually supportive unit within society. This is particularly true in less developed countries with little or no formal social security net. A poor outcome for one member of the household can be cushioned by the support of other members.

Income is reported within KIDS as a total monthly income and by a series of categories which can be utilised to generate a value for earned and unearned income. Unearned income is calculated as an average monthly income over the last 12 months. In the case of earned income, income received for the last month is reported. Therefore, earned income may be sensitive to when the survey took place within the particular household. Households who have had a “bad month” would report a much lower income than perhaps what may be reflective of their "normal" situation and vice versa. Given that there could be a multitude of reasons why income may differ from month to month there is no way to control for this.[206]

The approach taken by Wagstaff (2007) is utilised here: earned income is income which is generated from waged activities and also includes casual labour and bonuses, net income from agricultural production and family businesses and also any income received from renting out property or other assets. The accompanying KIDS codebook does not indicate specific sources of rental income and the range of income from this specific source is highly varied, suggesting a range of assets may be included.

Unearned income is defined within the context of this chapter as monies received from "non-labour sources", remittances and subsidies. There are three sources of subsidy within KIDS: transport, food and housing subsidies. Non-labour sources of income include payments such as pensions, unemployment insurance, disability and care grants, inheritance and interest/dividends.

In addition to reporting monthly household income (earned and unearned), households were asked to report the monthly income loss[207] associated with any shocks they experienced. The advantage of using a direct estimate, reported by each household, is that the determinants of household income do not need to be modelled. However, there are also a number of potential problems concerning the validity of such information. The first problem is that households were asked to indicate the income loss associated with a particular shock; some shocks are somewhat more precisely defined than others. For example, the "cut-off or decrease in remittances to the household" is relatively clearly defined in terms of income losses. The household knows how much they were receiving in remittances, and, following the shock, they know how much they are then in receipt of; the difference represents the income loss. The problem is that people may find it difficult to appropriately attribute an income loss directly to particular events, especially if multiple events occurred at the same time. If, for example, someone experienced a serious illness and ceased work, the household could be reporting the loss

[206] Ideally, information on income over a period of time such as a year would be more desirable as any such positive or negative swings in income would be smoothed out and reflect of the general level of income in the household.
[207] This is not differentiated as being from earned or unearned sources.

of their income plus any income loss from someone caring for that individual. In cases where two or more events occurred within a similar time period or across multiple individuals, it is more challenging to attribute losses to any one scenario. Therefore, households may misreport the true loss experienced. Moreover, as households are recalling information over a relatively lengthy period, their memory of the value of losses could be less reliable than would be desirable, which is what Bernard et al. (1984) refer to as "memory drift". In some cases it is possible to use tax records or other formal information to verify the reliability of such data (Atkinson and Micklewright, 1983) but such data is not available in this case and therefore there is no way to verify whether the income data reported is accurate.

6.1.2 Health shocks

Given that this chapter specifically considers health shocks, it is hypothesised using the Grossman (1972) model that individuals within a household who experience health shocks will experience a reduction in their stock of health capital. This loss of health capital could be relatively mild or severe, and is predicted within the model to lead to a loss of healthy days which are available for productive activity. Consequently, the potential opportunity for generating earned income in such cases will be lower, ceteris paribus. Additional falls in earned household income may also occur if other members of the household withdraw from the labour market or productive activities to provide care for the household member experiencing the shock. This chapter will also examine unearned income to see whether, during times of hardship, unearned income rises as a result of additional remittances (assuming that the members providing remittances are not also the sick members) which may be used to support the household, as outlined in *Chapter 2*.

As has been demonstrated in *Chapter 2*, health shocks can be defined in a number of different ways, such as: a deterioration in self-reported health status; acute and chronic hospitalisations; death of working-age individual; hospital stays longer than a week; reductions in body mass index; and major healthcare expenditures (Riphahn, 1999; Garcia-Gomez et al., 2010; Gustafsson-Wright et al., 2011; Wagstaff, 2007 and Gan et al., 2006). An examination of the KIDS survey highlights possible indicators that can be used to assess the impact of health shocks, including: incidence of sickness; general health status; expenditure on healthcare; and "economic shocks" (an addition to the original survey from 1998 onwards).

The "economic shocks" section in KIDS enables households to report on a range of "bad surprises" or events which have *"hurt the household financially over the past 5 years since 1993"*,[208] which are discussed in additional detail in *Chapter 3*. In particular, there are two indicators from this section of the KIDS survey which can be considered as representing health shocks: serious illness/injury and death.[209]

The health shocks reported within KIDS are reported by the household head for that household. Lee (2015) indicates that in such cases the household head is acting as a proxy respondent; usually this would be in the form of reporting information for each individual. The advantage of self-reported data is that they do not require physicians to assess individuals with tests which may be considered as invasive or for people to be measured (height/weight). As such it has become a very popular method of collecting data on population health as studies have shown that the information provided by respondents is a good predictor of mortality and health status (Idler and Kasl, 1995; Idler and Benyamini, 1997). A major disadvantage of the KIDS data is that health shocks are not self-reported for an individual within the economic shock module but are instead considered for the household collectively. Lee (2015) suggests that the accuracy of any information reported in this way is determined by the knowledge of the proxy respondent. Therefore the potential for reporting error is high if the proxy respondent does not have a good overview of the health of all household members.

The interpretation of the wording used within the survey is potentially sensitive and subjective. The use of the word "surprise" *in principle* implies that something is unexpected.[210] Whether a shock is truly unexpected is, however, very difficult to verify, as certainly in the case of someone experiencing long term or severe morbidity the

[208] This is amended to "Hurt the household financially over the past 6 years since 1998" when considering the 2004 dataset.
[209] Death is specifically listed as a shock in the 1998 survey module on shocks but not in 2004. Therefore a death variable has been created for the 2004 data.
[210] It should be noted that in practice the likelihood of a death being truly unexpected is questionable – one would expect that following a lengthy period of illness which is progressive households would anticipate death.

likelihood and expectation of death could be high. An accident such as a road traffic accident, in contrast, would be completely unexpected (Donovan et al., 2003). In practice, households may have some prior awareness of the *"bad surprise"* and could lead to them altering their behaviour in readiness for the event.

Moreover, the use of the term "serious" would also *in principle* imply that the morbidity experienced is not trivial and relatively minor. As has been outlined previously, people rate their health status given characteristics such as the level of education they possess and therefore it is possible that what might be considered "serious" for some may not be for others which could create a potential lack of consistency across households. A further disadvantage associated with this problem is that within KIDS this information is being reported by a household head rather than by specific individuals, and the responses could be conditioned by the view of the household head; some cases for some individuals could be considered to be more "serious" than others depending on an individual's relative position within the household.

As discussed previously, the lengthier the recall period is, the more difficult it may be for people to remember particular events or indeed the details about the event in question. The likelihood of a household recalling and subsequently reporting a relatively minor and inconsequential health shock which occurred within the last 5 or 6 years is rather unlikely as most are likely to have been forgotten. Rockwood (2015) describes the use of specific time periods as "time bounding"; the smaller the time period is, the more accurate the response. As the time period lengthens, the response is much more likely to be vaguer and less accurate. In contrast, a serious shock which has had repercussions for the household is much more likely to be recalled and reported and indeed for relatively uncommon events a longer reference period is required (Rockwood, 2015). Rockwood (2015) also indicates that the use of landmark events is also useful as they help to focus respondents on a particular period of time. Examples of landmark events are elections (and indeed elections took place when the KIDS survey was conducted) or a well-known event such as a natural disaster. It is important to note, however, that the validity of any reported information associated with each shock may be more questionable as the recall period becomes lengthier.

The Malawi IHS2 has a particularly interesting feature which is not available within KIDS but would be desirable: differentiation of the shock by the type of individual experiencing the shock. In particular, the Malawi IHS2 groups deaths into three categories: *"death of the household head, death of a working member of the household and death of other family member"*.[211] The aim of differentiating the shock would be to observe how the impact of a health shock is dependent upon the role of the individual within the household; this is, for example, a feature of the Yamano and Jayne (2004) study of Kenya. Health shocks experienced by non-productive members would have an indirect impact on household income if they then spent time caring for another household member rather than due to a direct loss of labour income. Unfortunately, it is not possible to examine shocks across different individuals given the data available. Therefore the measure of an injury or illness related health shock is defined here as a case where a household reports that they experienced *"serious injury or illness keeping household member from doing normal activities"* which must have occurred within the last 6 years in the 2004 survey and within the last 5 years for the 1998 survey.

A further disadvantage associated with the reported shock module is that households were only able to report one incidence of a shock. They were not asked if they had experienced multiple events which are very possible within a 5-6 year period. Nor were they asked to only mention the most serious shock. This applies to cases of both injury and illness and also death.[212] Therefore there are a number of possible scenarios that could have taken place which the household subsequently reports. The first is that households did experience only one incident of the serious injury/illness shock. The second is that they have reported only what they have deemed as the most "serious" shock, and the third is that they experienced multiple shocks. In this last case they could have then combined all of the shocks together, giving the date of the first shock and indicating that the shock has been ongoing to cover such a scenario. There is no way of knowing how the household has responded to this question and therefore it remains something to be aware of and consider within the discussion of the results which follow.

[211] Module AB: Recent shocks to household welfare.
[212] Deaths are only specifically reported as a shock in 1998 and therefore only recorded as one incident. In 2004 the information is drawn from the household roster. Therefore more than one incidence of death could be recorded in such a case.

At this point it is also appropriate to draw attention to some of the variation in how health shocks were reported between the 1998 and 2004 surveys. In 1998, households were asked to report details of when the shock occurred, the duration of the shock, the decrease in monthly income associated with the shock, the expenses arising from the shock and the value of any items lost where appropriate. Households were also asked about some of the coping strategies they may have used following the shock, such as whether they sold assets, borrowed from a lender, removed their children from school, used insurance or sought help from others. This information was not asked for in 2004. For this reason, no such comparison can be made in examining the coping strategies of households between 1998 and 2004 and is not explored within this chapter. Furthermore, in 1998 households reported a death shock within the bad economic shocks section. In 2004, this particular shock was removed from the list of potential shocks that the household could have experienced. Therefore a death shock[213] variable was created using information from the household roster for the 2004 case.

The reporting of a death is not subject to the same interpretation problems that may be associated with serious injury/illness and therefore has the advantage that *in principle* it is a much more clearly defined event. A death is an acute and final event resulting in the permanent loss of health capital to the household that cannot be replenished by further investment. Consequently, the death of particular household members also represents a permanent loss of income or productivity. As with serious injury/illness, a disadvantage of the KIDS data is that it is not clear who passed away.[214] The impact of the loss of an infant/child or elderly household member is likely to be different to the impact on the household following the loss of a working age member of the household. Referring back to the specific questioning used within the KIDS survey, the question proposed is that the event "financially hurt" the household and therefore all deaths could be reported. Consequently, it is not possible to exclude the possibility that children and the elderly are included within the 1998 death shock measure and due to the way deaths are reported in 2004,[215] they may be included.

[213] This may not be exactly comparable to the measure used for 1998 but given the information provided it is the next best measure available.
[214] In the 2004 survey, the gender of the deceased was reported.
[215] In 2004, household roster information was utilised.

6.1.3 Incidence and duration of shocks

Households reported whether they had experienced a shock of *"serious injury or illness keeping household member from doing normal activities"*. A health shock of this nature is not uncommon in the KwaZulu-Natal province, as Table 6.1.1 shows. In 1998, there were 241 households which experienced this particular shock, representing almost a fifth of all households. In 2004, this figure rises very slightly to 257 households representing 22 percent of *all* households.[216]

Table 6.1.1 Incidence of a serious injury or illness shock in 1998 and 2004 (all[217] households)

Shock status	Year		Total
	1998	*2004*	
Not-shocked	971	925	1,896
Shocked	241	257	498
Total	1,212	1,182	2,394

Source: KIDS (1998) and KIDS (2004)

Table 6.1.2 Duration of serious injury or illness shock in months (all households)[218]

	1998	2004
25th percentile	1	2.2
50th percentile	3	6
75th percentile	4	18
Mean	4.2	14.72
N	240	234
Missing	1	23

Source: KIDS (1998) and KIDS (2004)

As previously noted, the reported shock in KIDS is *"serious injury or illness"* and it is not possible to differentiate between them as there is not enough information available within the dataset to do this. However, it is possible to examine the duration of the reported illness or injury. The duration of a shock is conditioned on the shock arising because, by definition, if no shock is experienced then there is no duration of the shock. Table 6.1.2 shows the duration of serious injury or illness shocks affecting households in 1998 and 2004 conditioned on the shock occurring.

[216] The sample is reduced to 829 households when the identification of the household head is made for the regression estimations. In this case there are then 189 shocked households which represent 23 percent of the sample.
[217] This does not exclude those households where the household head was not readily identifiable.
[218] As in Table 6.1.1 households where the household head was not readily identifiable have not been excluded at this point.

In 1998, half of the shocked households reported that the shock lasted for three months or more. Five percent of the shocked households experienced a long-lasting shock of over a year. By 2004, the number of households experiencing more lengthy periods of serious injury or illness increased dramatically; half the households experience a shock lasting 12 months or more and a quarter experience a shock lasting over 18 months. The reported duration of the shocks between 1998 and 2004 are not directly comparable as in the 1998 survey people were asked to refer to a period of 5 years and this rose to 6 years for the 2004 survey. Therefore, the maximum permitted duration of the shock is longer in 2004 than in 1998. Putting aside the specific period of time that the shock may have occurred for, it is notable that in 2004, over two-fifths (44%) of households indicated that the shock had been ongoing, whilst a very slightly smaller proportion (39%) indicated that the shock had been ongoing in 1998.

Duration by itself may not be particularly informative; instead an approach taken by Donovan et al. (2003) which was outlined in *Chapter 2* is to use a specific period of morbidity. A health shock in the case of the Donovan et al. (2003) work is classified as someone who has stopped worked for a period of three months in the last year. Given that the KIDS dataset refers to a much longer period; 5 and 6 years respectively, I consider a slightly longer duration: six months as an indicator of a short lasting shock and all shocks lasting longer than six months are classified as long lasting. The benchmark case is that no shock has occurred at all[219].

6.1.4. Death shock

As in the case of a “serious injury or illness” outlined above in 1998 the survey also included a shock indicator for whether or not there was a “death of household member or other family member”. In 2004 this shock indicator was no longer available and so using information from the household roster outlined in *Chapter 3*, a comparable shock indicator was created[220]. However this measure only includes whether or not the

[219] The results from the analysis is not subsequently reported within section 6.3 as the use of duration of shocks does not add anything new in addition to the other results reported within the section.
[220] Within the household roster deaths were reported. All deaths were aggregated for each household. If the household experienced at least one death they were considered to have experienced a deathshock.

household experienced at least one death in the previous period, no other information is available as this measure has been constructed from the household roster information.

Death represents the ultimate health shock which may be the end result of a period of ill health or it could be the result of an accident (road traffic accident for example). As the indicator is presented as a binary variable with no additional detail as to the cause of death, it is not possible to separate out deaths resulting from serious illnesses from those caused by other means. This measure may also include deaths of elderly family members and children. The deaths of these household members may not be particularly influential in terms of impact upon household income as they would not be productive members of the household. The absence of such information is unfortunate as Yamano and Jayne (2004) indicate that the characteristics of the deceased can generate large differences in economic outcomes. In particular they find the death of a male head of household reduces off farm income by nearly 80 percent.

In order to be more confident that the deaths occurring within a household are more relevant to my estimation of health shocks, it would be useful to examine the age-death profile. If it were then found that the majority of deaths arising were of working age individuals, then I could be more satisfied that the use of such a measure is appropriate. However, whilst the household roster indicates whether a death has occurred in the preceding period, the respondent is not asked to disclose any details such as the individual's age in the household roster. Therefore, it is not possible to examine an age-death profile.

6.1.5 Fall in reported health status since previous election

As outlined in *Chapter 2*, there are a number of different ways in which we may examine health shocks' impact on economic outcomes. One approach used by Riphahn (1999) is to define a health shock as one where there is a sudden drop in self-reported health satisfaction. In this study, respondents were asked to rank their health satisfaction from a scale of 1-10 annually. Using this system, a fall in the reported ranking by 5 points or more is classified as a health shock.

The KIDS dataset has a comparable measure which may be utilised as an indicator of possible health shocks. Each individual who was classified as a core person reported how their health compares at the point of the survey, to how it did at the previous election[221]. Individuals who reported that their health was either "somewhat worse" or "much worse" than it had been previously are then coded as representing someone who has suffered a health shock. As it is possible to include age as an explanatory variable within the regression it is not necessary to have annual data. In *Chapter 2* it was observed that Yamano and Jayne (2004) and Jayne et al. (2005) both found that the economic outcomes that they explored were more pronounced when the shock was experienced by the household head.[222]

6.1.6 Additional variables

Given that the shock is experienced by a household as opposed to an individual, the indicators used within the estimations are based on the characteristics of the household head. For the 2004 data, a sub-sample of 829 households is considered, as the household head was not readily identifiable in the remaining cases.

For the purposes of the regression analysis, education has been categorised into five levels of education: no education, foundation, intermediate, senior and further education, as outlined in more detail in *Chapter 3*. Each level of education is included as a dummy with the exclusion of no education. As Table 6.1.3 shows, a large proportion of household heads have no education. Hence "no education" is a suitable base category for comparing the effect of differing levels of education upon household income in the presence of health shocks. In comparison to the IHS2 data, missing values for education are much less common: only one case in 1998 and 39 cases in 2004. In comparison, in the IHS2 data, information on education was missing in 24 percent of all cases, whereas in KIDS it is fewer than 5 percent.

[221] In the 1998 KIDS survey the relevant election was in 1994 and in 2004 the relevant election was in 1999.
[222] This indicator was used for household heads as a possible shock indicator which would enable me to control for the individual (the household head). The results from this however did not highlight anything particularly different in terms of the impact of the shock and therefore due to limitations of space the results from the use of this indicator as a health shock are not presented within the subsequent analysis.

Levels of education have been chosen as opposed to the number of years of education as the achievement of a level of education may be more informative. As previously noted, one could spend a long time in school without achieving a particular educational qualification.

Table 6.1.3 Education levels of household heads in 1998 and 2004

Level of education	1998		2004	
	Number	Percentage	Number	Percentage
Missing	1	0.08	39	4.70
No education	289	23.84	172	20.75
Foundation	283	23.35	195	23.52
Intermediate	300	24.75	209	25.21
Senior	212	17.49	106	12.79
Further	127	10.48	108	13.03
Total	1,212	100.00	829	100.00

Source: KIDS (1998) and KIDS (2004)

Indicators for the gender of the household head, household size and the number of dependents[223] have also been included within the estimations. In addition, the type of employment sector that the household head is in may also have some bearing on the likelihood of labour supply adjustment. Households engaged in farming activities are more likely to be able to adjust labour supply compared to those households who are engaged in other off-farm activities. Hence, Wagstaff (2007) suggests that rural households may potentially be more protected from the negative effects of a health shock than households which are engaged in other activities within the urban area.

The KIDS dataset holds information on the type of employment individuals are engaged in which is outlined in detail in *Chapter 3*. The results presented consider a series of dummies to represent broad types of employment status; regular employment, casual employment, self-employment, not in the labour pool[224] and unemployed. This is a slightly more detailed approach than that of Gustafsson-Wright et al. (2011), who differentiate between those in employment and those who are not. However, the approach used is not as detailed as Riphahn (1999) and Garcia-Gomez et al. (2010). Each employment activity is included as a dummy variable with the exclusion of unemployment.

[223] Number of children aged less than 15 years old in this case but it could also be all dependents, i.e. children and elderly relatives.

[224] This category includes housewives, those in education, retirees and those who answered “other”.

6.2 Estimation results: the impact of a health shock on household income

Table 6.2.1 provides a summary of the variables used within the subsequent analysis and Tables 6.2.2-7 present descriptive statistics for the 1998 and 2004 datasets. It should be noted that earned income can be negative as it includes the net income from agricultural activities and businesses; these can be loss makers.

Table 6.2.1 Summary list of variables

Variable	Variable definition
HHsize	number of household members
HHchild	number of children in household aged less than 15
HHgender	dummy: equal to 1 if household head is female
HHage	age of household head
HHagesq	age head of household squared
No educ	dummy: equal to 1 if household head has no education
Foundation	dummy: equal to 1 if household head has foundation level education
Intermediate	dummy: equal to 1 if household head has intermediate level education
Senior	dummy: equal to 1 if household head has senior level education
FET	dummy: equal to 1 if household head has further level education or training
Regular	dummy: equal to 1 if household head is engaged in regular waged employment
Casual	dummy: equal to 1 if household head is engaged in casual waged employment
Selfemployed	dummy: equal to 1 if household head is self-employed
NILP	dummy: equal to 1 if household head is not active in the labour market
Unemployed	dummy: equal to 1 if household head is unemployed
Other	dummy: equal to 1 if household head indicates that their employment falls into none of the other categories
Illshock	dummy: equal to 1 if the household has experienced a serious injury or illness since the last survey
Deathshock	dummy: equal to 1 if the household has experienced a death since the last survey
HH earned income	Monthly household income from earned sources in South African Rand
HH unearned income	Monthly household income from non-earned sources in South African Rand
Log (HH earned income)	The natural logarithm of monthly household income from earned sources
Log (HH unearned income)	The natural logarithm of monthly household income from non-earned sources

A brief examination of Tables 6.2.2-7 highlights some similarities and differences both between the groups, shocked and non-shocked, and also across the two time periods. The first notable point in all cases (both by shock status and year) is that the mean household income (both earned and unearned) is being driven upwards by some very large values of income. Moreover, it is notable that the earned income reported by non-shocked households is somewhat larger than that reported by shocked households and this is true for both the 1998 and the 2004 data. This is not, however, surprising as the Grossman (1972) model predicts that healthy households will have more "healthy days" available for productive activity and therefore income from earned sources would be higher in such cases, everything else being equal.

Table 6.2.2 Descriptive statistics 1998 - all households

	1998 – All households						
	Mean	Median	Min	Max	Standard deviation	N	Missing
HHsize	7.29	6	1	34	2.68	1212	0
HHadults	4.76	4	1	24	2.68	1212	0
HHchild	2.53	2	0	15	2.37	1212	0
HHgender	0.60	1	0	1	0.49	1212	0
HHage	53.12	52	19	100	14.16	1212	0
HHagesq	3022.60	2704	361	10000	1576.04	1212	0
No educ	0.24	0	0	1	0.43	1211	1
Foundation	0.23	0	0	1	0.42	1211	1
Intermediate	0.25	0	0	1	0.43	1211	1
Senior	0.18	0	0	1	0.38	1211	1
FET	0.10	0	0	1	0.31	1211	1
Regular	0.40	0	0	1	0.49	1212	0
Casual	0.07	0	0	1	0.25	1212	0
Selfemployed	0.06	0	0	1	0.24	1212	0
NILP	0.35	0	0	1	0.48	1212	0
Unemployed	0.11	0	0	1	0.31	1212	0
Other	0.01	0	0	1	0.09	1212	0
Illshock	0.20	0	0	1	0.40	1212	0
Deathshock	0.35	0	0	1	0.47	1212	0
Log (HH earned income)	6.80	6.98	0.08	10.62	1.58	1141	71
Log (HH unearned income)	5.28	5.51	0.51	8.29	1.14	430	782

Source: KIDS (1998)

Table 6.2.3Descriptive statistics 1998 – non shocked only[225]

	1998 – Non shocked households						
	Mean	Median	Min	Max	Standard deviation	N	Missing
HHsize	7.00	6	1	28	4.21	971	0
HHadults	4.58	4	1	15	2.54	971	0
HHchild	2.43	2	0	15	2.30	971	0
HHgender	0.59	1	0	1	0.49	971	0
HHage	52.55	51	19	95	14.06	971	0
HHagesq	2959.65	2601	361	9025	1553.50	971	0
No educ	0.23	0	0	1	0.42	970	1
Foundation	0.23	0	0	1	0.42	970	1
Intermediate	0.25	0	0	1	0.43	970	1
Senior	0.18	0	0	1	0.39	970	1
FET	0.11	0	0	1	0.32	970	1
Regular	0.42	0	0	1	0.49	971	0
Casual	0.06	0	0	1	0.25	971	0
Selfemployed	0.06	0	0	1	0.24	971	0
NILP	0.34	0	0	1	0.47	971	0
Unemployed	0.11	0	0	1	0.31	971	0
Other	0.01	0	0	1	0.10	971	0
Log (HH earned income)	6.83	7.08	0.08	10.62	1.67	906	65
Log (HH unearned income)	5.27	5.52	0.51	8.29	1.15	319	652

Source: KIDS (1998)

[225] Based on shock being related to "serious injury or illness".

Table 6.2.4 Descriptive statistics for all 1998 – shocked only[226]

	1998 – Non shocked households						
	Mean	Median	Min	Max	Standard deviation	N	Missing
HHsize	8.43	7	1	34	4.97	241	0
HHadults	5.51	5	1	24	3.09	241	0
HHchild	2.92	2	0	13	2.59	241	0
HHgender	0.61	1	0	1	0.49	241	0
HHage	55.42	55	24	100	14.35	241	0
HHagesq	3276.23	3025	576	10000	1642.76	241	0
No educ	0.29	0	0	1	0.45	241	0
Foundation	0.25	0	0	1	0.44	241	0
Intermediate	0.25	0	0	1	0.43	241	0
Senior	0.14	0	0	1	0.35	241	0
FET	0.07	0	0	1	0.26	241	0
Regular	0.32	0	0	1	0.47	241	0
Casual	0.07	0	0	1	0.26	241	0
Selfemployed	0.07	0	0	1	0.26	241	0
NILP	0.41	0	0	1	0.49	241	0
Unemployed	0.12	0	0	1	0.33	241	0
Other	0	0	0	0	0	241	0
Mduration	4.2	3	0	40	5.76	240	1
Log (HH earned income)	6.65	6.70	1.47	9.55	1.50	235	6
Log (HH unearned income)	5.30	5.42	0.92	7.89	1.10	111	130

Source: KIDS (1998)

[226] Based on shock being related to "serious injury or illness".

Table 6.2.5 Descriptive statistics 2004 –all households

	2004 – All households						
	Mean	Median	Min	Max	Standard deviation	N	Missing
HHsize	7.31	7	1	31	4.33	829	0
HHadults	4.69	4	1	19	2.58	829	0
HHchild	2.61	2	0	14	2.38	829	0
HHgender	0.45	1	0	1	0.49	829	0
HHage	57.72	57	19	90	12.94	761	68
HHagesq	3498.72	3249	361	9604	1523.34	761	68
No educ	0.22	0	0	1	0.41	790	39
Foundation	0.25	0	0	1	0.43	790	39
Intermediate	0.26	0	0	1	0.44	790	39
Senior	0.13	0	0	1	0.34	790	39
FET	0.14	0	0	1	0.34	790	39
Regular	0.21	0	0	1	0.41	771	58
Casual	0.08	0	0	1	0.28	771	58
Selfemployed	0.08	0	0	1	0.27	771	58
NILP	0.44	0	0	1	0.50	771	58
Unemployed	0.17	0	0	1	0.38	771	58
Other	0.01	0	0	1	0.12	771	58
Illshock	0.23	0	0	1	0.42	817	12
Deathsock	0.39	0	0	1	0.48	829	0
Log (HH earned income)	6.886	6.903	-10.102	11.429	1.712	790	39
Log (HH unearned income)	6.661	6.677	2.120	11.209	0.942	653	176

Source: KIDS (2004)

Table 6.2.6 Descriptive statistics 2004 – non shocked households

	2004 – Non shocked households						
	Mean	Median	Min	Max	Standard deviation	N	Missing
HHsize	7.06	6	1	31	4.33	628	0
HHadults	4.53	4	1	19	2.59	628	0
HHchild	2.53	2	0	14	2.37	628	0
HHgender	0.44	0	0	1	0.50	628	0
HHage	57.49	56.50	23	98	13.03	574	54
HHagesq	3474.42	3192.50	529	9604	1543.04	574	54
No educ	0.21	0	0	1	0.41	599	29
Foundation	0.25	0	0	1	0.43	599	29
Intermediate	0.25	0	0	1	0.44	599	29
Senior	0.14	0	0	1	0.35	599	29
FET	0.15	0	0	1	0.35	599	29
Regular	0.23	0	0	1	0.42	582	46
Casual	0.08	0	0	1	0.27	582	46
Selfemployed	0.09	0	0	1	0.29	582	46
NILP	0.43	0	0	1	0.49	582	46
Unemployed	0.16	0	0	1	0.37	582	46
Other	0.02	0	0	1	0.13	582	46
Log (HH earned income)	6.931	6.947	-10.102	11.429	1.793	607	21
Log (HH unearned income)	6.668	6.646	2.996	11.209	0.938	475	153

Source: KIDS (2004)

Table 6.2.7 Descriptive statistics 2004 – shocked households

	2004 – Shocked households						
	Mean	Median	Min	Max	Standard deviation	N	Missing
HHsize	7.95	7	1	24	4.20	189	0
HHadults	5.14	5	1	14	2.45	189	0
HHchild	2.80	2	0	11	2.39	189	0
HHgender	0.47	0	0	1	0.50	189	0
HHage	59.29	59	35	86	11.97	175	14
HHagesq	3685.65	3481	1225	7396	1440.05	175	14
No educ	0.25	0	0	1	0.44	179	10
Foundation	0.23	0	0	1	0.42	179	10
Intermediate	0.31	0	0	1	0.46	179	10
Senior	0.09	0	0	1	0.29	179	10
FET	0.11	0	0	1	0.31	179	10
Regular	0.14	0	0	1	0.35	177	12
Casual	0.10	0	0	1	0.30	177	12
Selfemployed	0.06	0	0	1	0.23	177	12
NILP	0.49	0	0	1	0.50	177	12
Unemployed	0.21	0	0	1	0.41	177	12
Other	0.01	0	0	1	0.08	177	12
Duration	14.21	6	0	75	18.46	171	18
Log (HH earned income)	6.736	6.776	2.773	9.422	1.402	183	6
Log (HH unearned income)	6.665	6.813	2.120	9.972	6.666	170	19

Source: KIDS (2004)

Unearned income, which comes from sources such as subsidies for food, transportation and housing, is in contrast higher amongst shocked households across both time periods; this may be reflective of additional support which is given to more vulnerable households. Looking specifically at the key source of the difference in unearned income across the two groups; shocked and non-shocked, indicates that it is subsidies that the household is in receipt of rather than remittances that led to shocked households having greater unearned income relative to non-shocked households.

Households experiencing a serious injury/illness shock tend to be headed by slightly older individuals than non-shocked households by around 2-3 years in both time periods. The gender of the household head appears to be broadly similar across both types of household and does not suggest that shocks are more commonly experienced by female headed households, although the proportion of female headed households is much greater in 1998 compared to 2004. However, as previously noted, the identification of the household head was somewhat challenging in 2004 and so some households are not included within the analysis where a clear head was not identifiable. Household size and other measures of composition such as the number of children living within a household are comparable over time and across household types.

The level of education achieved by the household head varies slightly across the household types: slightly larger proportions of shocked households are headed by a head with no formal education and generally the level of education held by a household head of shocked households is lower in comparison to non-shocked households. Shocked households also tend to have a higher proportion of individuals who are not in the labour pool; this includes retirees, housewives, the disabled and those who are in full time education; this is true for households in 1998 and 2004. A slight difference that arises over time is the proportion of households that are headed by individuals who are unemployed; in 1998 the proportion of unemployed household heads is similar across shocked and non-shocked households. In 2004, the proportion of households headed by an unemployed individual is much higher amongst shocked households compared to non-shocked households.

Table 6.2.2 shows that around one fifth of households in 1998 experienced a serious injury/illness shock and just over a third experienced a death shock. These figures increase very slightly in the 2004 dataset but as the 2004 data considers an additional year; this could explain the slight increase. In addition, by 2004 people are older and therefore the stock of health capital would naturally be decreasing and such incidents may be more likely. What is also clear from Table 6.2.8 is that households that have been shocked by a serious injury/illness shock in 1998 are also more likely to experience this shock in future. Going back to the Grossman (1972) model, it is apparent that shocked households receive lower earned incomes on average relative to non-shocked households. Ceteris paribus, such households are then more constrained in the next period as they have less income to purchase health inputs such as better nutrition which would increase the stock of health capital. In the presence of initial constraints on income, households may be constrained in the next period in terms of undertaking investments in health capital and may be more vulnerable to further shocks and potentially more challenging shocks in the next period.

A direct comparison of the duration of a serious injury/illness shock is not possible as the time periods referred to in the two surveys differ but it is clear that the average duration of a shock (injury/illness) does increase over the survey periods. Only a small number of households[227] actually experienced a shock which lasted over 60 months in the 2004 data. If these cases are excluded to compare the duration of the shock across the two surveys, it is observed that the mean duration of a serious injury/illness shock in 2004 remains very high at 11.48 months, over double the duration experienced in 1998.

Tables 6.2.2-7 highlight some similarities and differences within the 1998 and 2004 KIDS data. Before proceeding, it is also important to consider whether there are any particular characteristics of the households, which make them more likely to experience health shocks as the literature indicates that in certain circumstances the presence of health challenges may be highly correlated with individuals' characteristics. Taking HIV/AIDS as one possible health shock, Pitayanon et al. (1997) indicate that in a 1995 report for the Ministry of Public Health, more than 60 percent of all cases of HIV/AIDS reported in Thailand are amongst labourers and agricultural workers. These individuals are generally poor and have the least amount of education.

In Uganda, disruption in the late 1970s and early 1980s in the region led to hardship for many and led to some vulnerable women exchanging sexual favours for cash, particularly with men who were more educated and more wealthy who were travelling along the main transport routes in the area (Barnett and Whiteside, 2006). Given that the KIDS survey is a panel, it is possible, to assess whether the characteristics of the household in 1998 offer any indication of a health shock occurring in 2004. A similar model is not possible for the 1998 data as the economic shocks module did not exist in the 1993 survey; it was only introduced in 1998.

Using a probit estimator one can examine whether there are certain household characteristics that are associated with the probability of experiencing future health shocks as shown in Table 6.2.8. A Wald F-test fails to reject the null hypothesis that all the coefficients are jointly equal to zero (column 1).

---

[227] 38 households.

Table 6.2.8 Probit estimation results for likelihood of household[228] experiencing a future health shock.

	(1)	(2)
Log(HH earned income)	-0.058 (0.038)	-0.056 (0.038)
HHchilld	-0.054 (0.024)**	-0.052 (0.024)**
HHadults	0.023 (0.021)	-0.014 (0.022)
HHgender	0.196 (0.118)	0.177 (0.119)
HHage	-0.005 (0.026)	-0.003 (0.027)
HHagesq	0.0001 (0.0002)	0.0001 (0.0002)
Foundation	-0.062 (0.153)	-0.068 (0.015)
Intermediate	-0.059 (0.159)	-0.062 (0.159)
Senior	-0.128 (0.197)	-0.123 (0.197)
FET	-0.084 (0.245)	-0.080 (0.243)
Regular	-0.130 (0.172)	-0.126 (0.172)
Casual	-0.059 (0.246)	-0.061 (0.249)
Self-employed	-0.062 (0.239)	-0.088 (0.242)
NILP	-0.168 (0.199)	-0.714 (0.200)
Other	0.287 (0.505)	0.377 (0.505)
Illshock		0.331 (0.119)***
Wald chi2	$\chi^2(15) = 20.91$	$\chi^2(16) = 28.49$
N	728	728

Source: KIDS. Notes: *** significant at 1% level, ** significant at 5% level, * significant at 10% level; robust standard errors in parentheses

If an indicator of whether the household experienced an illness health shock between 1993 and 1998 is also included, as shown in Table 6.2.8, column (2), it is found that experiencing an illness related health shock in 1998 increases the probability of experiencing the same type of shock in 2004.[229] In sum, previous shocks appear to increase the likelihood of households experiencing a future health shock.[230]

The implication of this is twofold. Firstly, households which have previously been shocked are more likely to experience future shocks, ceteris paribus. Secondly, if the

[228] Characteristics of households in 1998 and a shock occurring in 2004.
[229] The marginal effect of being shocked in 1998 is to increase the probability of being shocked again in 2004 by 11 percentage points. Please note marginal effects are not reported within Table 6.2.8.
[230] Specifically an illness or injury related shock.

impact of health shocks upon household income is large, the ability of the household to cope with future shocks could be highly constrained, which is a great concern if they are also more likely to experience future shocks. Furthermore, households which have been previously shocked may be more likely to change their behaviour ex ante relative to households which have not previously been shocked. This is an issue which could be explored in future research assuming the "shock" module is retained within the KIDS survey.

6.3. Regression analysis

Using the definition of income outlined in section 6.1, the dependent variable used in the estimated models is monthly earned household income in natural logarithms.[231] A log transformation has been selected due the data being highly skewed as shown in Figure 6.3.1.As an alternative monthly unearned income is also used within some of the analysis.

The full list of variables used in the regression analysis is presented in Table 6.2.1. Table 6.3.1 compares the nominal monthly earned income between households that have experienced a health shock; “shocked” households and non-shocked (healthy) households in 1998 and 2004. An examination of the unconditional means indicates that the nominal monthly earned household income of shocked households is much lower relative to non-shocked households in both 1998 and 2004. However, we cannot immediately conclude that this is a causal impact of shocks on income for two reasons. The first is the problem of confounding variables: households that are shocked may have lower incomes for other, observable reasons. The second is the problem of reverse causality: lower household income may lead to a greater likelihood of suffering a shock. The estimations which follow attempt to address these problems.

[231] Given the relatively small number of missing values for income as a result of taking logs, an alternative transformation such as the inverse hyperbolic sine has not been undertaken.

Figure 6.3.1 Distribution[232] of earned monthly nominal income 1998 and 2004

1998 *2004*

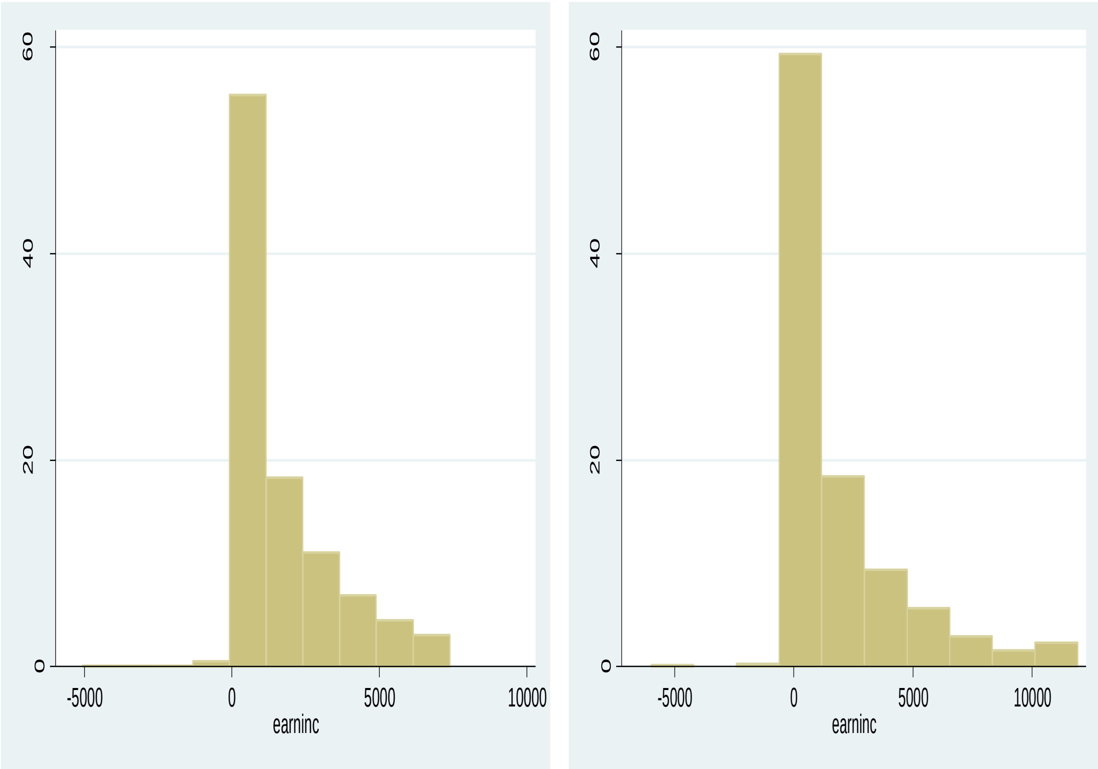


Table 6.3.1 Monthly (nominal) earned income of households in 1998 and 2004 relative to shock status in South African Rand[233]

	1998		2004	
Shock status	Shocked	Non-shocked	Shocked	Non-shocked
Median	785.42	1083.33	866.67	981.89
Mean	1688.76	2212.80	1825.66	3421.82

Source: KIDS (1998) and KIDS (2004)

Using the framework outlined in section 6.1, the models that follow seek to estimate the effect of health shocks on household income. A shock to the health of an individual reduces their stock of health capital which is similar to other forms of human capital; additional stock of such capital enables people to be more productive.

I first report a series of OLS regressions in which I control for a number of observable characteristics of households. Healthy households, ceteris paribus, will have a greater

[232] Excluding the top five percent of households, who reported a nominal monthly earned income in excess of 7,390 Rand in 1998; 11,917 Rand in 2004

[233] To give an indication of the Rand value in UK Sterling, in 1998 100 Rand was approximately £11 and in 2004, the approximate value of 100 Rand was £9.

stock of human capital relative to shocked households and are therefore predicted to generate greater incomes controlling for other household characteristics which are proxied by the characteristics of the household head. The coefficient on the shock may be interpreted as the average treatment effect of the shock on household income (which may be earned or unearned given the specification), ceteris paribus. The results in Table 6.3.2 show that one cannot reliably estimate the effect of a serious injury or illness shock on earned household incomes. This is shown by the large standard error of the shock in both the 1998 and 2004 estimations. In 1998, the 95% confidence interval for "illshock" indicates that households experiencing a shock could have an earned monthly income which is estimated to be between 30 percent lower and 10 percent higher than a healthy household, ceteris paribus. This very large range means we cannot makes a precise statement about the size of the average treatment effect of such a health shock.

With the exception of the coefficient on foundation education in 2004, the coefficients on the differing levels of education are all statistically significant at the 1% level and demonstrate that in households where the household head possesses at least some education,[234] there are large positive returns to household income. Whilst there are more categories for the KIDS data relative to the IHS2, the results are broadly similar in that foundation/primary level education generates large returns relative to having no education at all and the returns to further education are extremely large. In 1998 the coefficient on gender shows that male headed households will have an earned income that is almost 19% more than their comparative female headed households and this is significant at the 5% significance level. In 2004, the coefficient on gender is somewhat unexpected as it is negative; however the coefficient is not statistically different from zero. In both 1998 and 2004 age and age-squared have the anticipated signs, showing that earned household income initially increases with the age of the household head and then declines, consistent with human capital theory literature.

Additional adult household members have a positive effect on household income; increasing monthly earned household income by 9.2% in 1998 and by 9.4% in 2004.

[234] No education is the omitted category.

Table 6.3.2 OLS estimates of the impact of a "serious injury or illness" health shock on monthly earned household income

Dependent Variable: Natural Logarithm of Earned Monthly Household Income		
	1998 (1)	2004 (2)
Illshock	-0.096 (0.101)	-0.076 (0.120)
HHadults	0.093 (0.019)***	0.095 (0.026)***
HHchildren	-0.051 (0.020)**	-0.096 (0.029)***
HHgender	0.190 (0.089)**	-0.106 (0.122)
Hage	0.040 (0.021)*	0.061 (0.035)*
Hagesq	-0.0002 (0.0002)	-0.005 (0.0003)
Foundation	0.457 (0.126)***	0.154 (0.147)
Intermediate	0.882 (0.126)***	0.622 (0.159)***
Senior	1.612 (0.132)***	0.948 (0.294)***
FET	2.312 (0.130)***	1.260 (0.203)***
Regular	0.834 (0.148)***	1.564 (0.163)***
Casual	0.281 (0.207)	0.672 (0.182)***
Self employed	0.840 (0.207)***	1.157 (0.241)***
NILP	-0.038 (0.183)	0.227 (0.193)
Other	-1.176 (0.438)***	-0.485 (0.416)
Constant	3.503 (0.592)	3.794 (1.044)
F	F(15,1124)=47.19 (p-value = 0.00)	F(15,68) = 28.23 (p-value = 0.00)
R^2	0.2957	0.2736
N	1140	697

Source: KIDS. Notes: *** significant at 1% level **significant at the 5% level, * significant at the 10% level; robust standard errors in parenthesis

This is not surprising, as additional adults increase the productive capacity of the household. As anticipated the number of children in the household reduces household income, in 1998 by 5% and 19% in 2004; households with more children will require household members to stay at home and care for the children, which means that fewer members will be in the labour pool earning income or engaged in productive activities. The signs on the coefficients related to the type of employment that the household head is engaged in are as expected.

Table 6.3.3 OLS estimates of the impact of a "death" health shock on monthly earned household income

Dependent Variable: Natural Logarithm of Earned Monthly Household Income		
	1998 (1)	2004 (2)
Deathshock	0.046 (0.083)	-0.202 (0.124)*
HHadults	0.091 (0.019)***	0.095 (0.026)***
HHchild	-0.051 (0.020)**	-0.089 (0.030)***
HHgender	0.191 (0.090)**	-0.073 (0.120)
Hage	0.040 (0.021)**	0.065 (0.035)*
Hagesq	-0.0002 (0.0002)	-0.0005 (0.0003)*
Foundation	0.453 (0.126)***	0.161 (0.147)
Intermediate	0.879 (0.126)***	0.632 (0.158)***
Senior	1.618 (0.132)***	0.916 (0.291)***
FET	2.315 (0.130)***	1.255 (0.203)***
Regular	0.837 (0.147)***	1.570 (0.161)***
Casual	0.276 (0.205)	0.657 (0.181)***
Self-employed	0.837 (0.207)***	1.154 (0.239)***
NILP	-0.039 (0.183)	0.224 (0.193)
Other	-1.153 (0.440)***	0.453 (0.407)
Constant	3.473 (0.593)	3.704 (1.041)
F	F(15,1124) = 46.40 (p-value = 0.00)	F(15,681) = 27.80 (p-value = …)
R^2	0.2953	0.2763
N	1140	697

Source: KIDS. Notes: *** significant at 1% level **significant at the 5% level, * significant at the 10% level; robust standard errors in parenthesis

Households which have a head engaged in regular waged activity, earn substantially more than those which are headed by those where the head is unemployed. I reject the null hypothesis that the effect of a head being engaged in regular waged employment has the same effect upon household earned income as being self-employed, in 2004. However, the results from 1998 suggest that I cannot reject the null hypothesis that the effect of the household head being employed on a regular wage basis or self-employed have an equal effect upon household monthly income.

If a health shock in the extreme is now considered, a death, the observed coefficients for 1998 are fairly similar to those reported for a serious injury or illness shock. Education of the household head is highly statistically significant at most levels, increasing monthly earned household income by large amounts relative to households which are headed by those with no formal education, indicating as noted previously, substantial returns to education.

An examination of the coefficients on different types of employment status also shows a similar pattern to that observed in Table 6.3.2; households that have a head engaged in regular and self-employment earn substantially more than those households where the head is unemployed. The coefficient on gender is also statistically significant at the 5% level in 1998, indicating that male headed households earn more than their female headed equivalents.

The results presented in Table 6.3.2 indicate a very imprecise estimation of the effect of a death shock upon households in 1998 as was observed previously. In this case, households experiencing a death shock may experience monthly earned incomes which range between 12 percent less than healthy households and up to 21 percent more, ceteris paribus. As noted previously, this result does not provide any reasonable indication as to what the real impact of a death shock may be. The results for 2004 are also very similar with a very broad confidence interval which again renders the estimated effect of a death shock impossible to state with any precision. In 2004, the estimated impact of a death shock indicates that shocked households could receive earned incomes which range between 45 percent lower and 4 percent higher than those received by healthy households, ceteris paribus.

In the results presented above, nominal monthly earned income has been used. The next set of results consider nominal monthly unearned income. In *Chapter 2*, it was observed that earned and unearned[235] income behave in different ways in the presence of morbidity. Earned income will tend to decline as household members withdraw from the labour market whereas unearned income tends to increase as other (non-resident) members of the household increase their assistance to the household in times of hardship. Therefore, it is hypothesised that health shocks will increase household unearned income.

[235] For example gifts and remittances.

Table 6.3.4 OLS estimates of the impact of health shocks on monthly unearned household income in 1998 and 2004

	Dependent Variable: Natural Logarithm of Unearned Monthly Household Income			
	1998		*2004*	
	Illshock (1)	Death (2)	Illshock (3)	Death (4)
Healthshock	0.768 (0.189)***	-0.310 (0.169)*	0.767 (0.211)***	-0.096 (0.206)
HHadults	0.176 (0.035)***	0.191 (0.034)***	0.117 (0.044)***	0.109 (0.043)**
HHchild	0.006 (0.038)	0.008 (0.039)	0.216 (0.048)***	0.230 (0.047)***
HHgender	0.204 (0.174)	0.201 (0.174)	0.238 (0.214)	0.209 (0.222)
Hage	-0.035 (0.036)	-0.036 (0.037)	0.013 (0.059)	0.027 (0.062)
Hagesq	0.0006 (0.0003)*	0.0006 (0.0003)*	0.0002 (0.0005)	0.0001 (0.0005)
Foundation	0.331 (0.217)	0.350 (0.218)	0.281 (0.248)	0.262 (0.248)
Intermediate	0.193 (0.229)	0.209 (0.229)	0.176 (0.269)	0.192 (0.271)
Senior	-0.321 (0.292)	-0.328 (0.295)	0.035 (0.390)	-0.015 (0.391)
FET	0.766 (0.397)	0.768 (0.401)*	1.052 (0.397)***	1.009 (0.396)***
Regular	-0.717 (0.314)**	-0.736 (0.319)**	-0.560 (0.404)	-0.731 (0.403)*
Casual	-1.690 (0.432)***	-1.653 (0.436)***	0.181 (0.468)	0.145 (0.473)
Self-employed	-1.088 (0.486)**	-1.072 (0.491)**	-0.190 (0.501)	-0.427 (0.491)
NILP	1.820 (0.330)***	1.834 (0.334)***	1.624 (0.382)***	1.507 (0.381)***
Other	1.799 (0.847)*	1.623 (0.862)*	0.919 (1.000)	0.707 (1.023)
Constant	3.353 (1.051)	3.554 (1.066)	2.018 (1.793)	1.855 (1.864)
F	F(15,1154) = 52.99 (p-value = 0.00)	F(15,1154) = 51.82 (p-value = 0.00)	F(15,705) = 22.15 (p-value = 0.00)	F(15,717) = 22.12 (p-value = 0.00)
R^2	0.3016	0.2950	0.2789	0.2654
N	1170	1170	721	733

Source: KIDS Notes: *** significant at 1% level **significant at the 5% level, * significant at the 10% level; robust standard errors in parentheses

Table 6.3.4 presents results using nominal monthly unearned income as the dependent variable. A relatively large proportion of households report that their unearned income is zero as they don't receive any and therefore using a standard log transformation would lead to a substantial drop in the number of observations.

In the case of earned income this was not a particular concern as a relatively large number of observations are still utilised. However, in the case of unearned income, in 1998, this would lead to a large reduction in the sample. Therefore, the results presented for unearned income utilise the inverse hyperbolic sine transformation[236] to maximise the number of cases given the large number of households reporting that they receive no unearned income.

In the case of unearned income reported in Table 6.3.4, the coefficient on the shock variable is positive and very large, in the case of an illness. Even at the lower end of the 95% confidence interval, experiencing an illness shock increases unearned income by over 30 percent in both 1998 and 2004. This supports the theory previously considered: unearned income increases following a shock.

Households who are headed by those in regular, casual or self-employment receive much less unearned income relative to households headed by someone who is unemployed, ceteris paribus. Practically, households which are able to earn income may be perceived (and indeed be) less vulnerable than households which are not able to earn income; instead, such households seek support from other sources. At a practical level, the employment coefficients are also very imprecisely measured due to the large standard errors and therefore, one cannot reliably conclude the true impact of employment status as in earlier parts of the discussion.

6.4 Propensity score matching (PSM)

In experimental data a control group can be clearly defined as a baseline for comparison. Effectively one can observe the outcome for the treated and the untreated. An examination of the difference between the two outcomes based on whether the treatment occurred or not (non-treatment group) then provides an average treatment effect.

The difficulty with non-experimental data is that it is impossible to know what the outcome would have been in the absence of the treatment and vice versa. Effectively there is not a counterfactual case: what would household income have been in the

[236] Defined as $\log(y_i+(y_i^2+1)^{1/2})$. Unless the value of y is very small, this transformation is approximately equal to $\log(2y_i)$ or $\log(2)+\log(y_i)$. The coefficient on illshock can therefore be interpreted as an elasticity in much the same way as it would be in a standard log transformation. http://worthwhile.typepad.com/worthwhile_canadian_initi/2011/07/a-rant-on-inverse-hyperbolic-sine-transformations.html

absence of a shock for shocked households and vice versa. Given that the KIDS data is not experimental, the inclusion of a household into one group or the other may not be random; instead it could be that there are characteristics of households (selection bias) that determine whether it is shocked or not and these characteristics also determine the outcome (Caliendo and Kopeinig, 2005). In the first part of this chapter, I addressed this issue using the method of OLS and including controls for observable characteristics of households, but this led to results with very imprecise estimates of the treatment effect. Here, I use a different method, matching, to control for differences in the characteristics of shocked and unshocked households and see if it leads to more precise estimates of the treatment effect. An advantage of PSM relative to OLS is that the two groups don't have to be similar across all covariates. PSM reduces the characteristics down to one dimension and then examines the treatment effect based on the propensity score[237].

For example, an individual with no education may have a relatively limited knowledge about disease prevention, increasing the likelihood of them experiencing morbidity. Moreover, that same individual has a smaller stock of human capital and would therefore be expected according to the human capital literature to receive lower wages in the labour market. Comparing this individual with someone who has not experienced a shock may not compare like-with-like. An individual not experiencing morbidity may be more educated and hence avoid activities which make them more likely to experience morbidity *and* as they are more educated they will also receive higher wages. The two individuals in this example are fundamentally different and would be expected to have different wages even without accounting for any morbidity (the treatment).

Therefore, there is a requirement to create two groups that are similar enough in their characteristics to one another so that a comparison between the groups to be made in order to estimate the effect of the "treatment" (Dehejia and Wahba, 2002). In principle if two groups only had one characteristic that separated them one would be able to assign cases into two groups fairly easily; treated and non-treated, where they only differ by the treatment effect (Dehejia and Wahba, 2002). In practice there are a number of characteristics that describe an individual or in the case of the KIDS data, a

[237] http://fmwww.bc.edu/EC-C/S2013/823/EC823.S2013.nn12.slides.pdf

household. Therefore, rather than trying to select and then match cases on a specific singular characteristic, score matching may be used. The approach taken within this part of the chapter is to specifically consider propensity score matching (PSM). This method effectively assigns a score to the different households based on the probability that the household in question will receive the treatment (experience a health shock) given their observable characteristics and seeks to mimic random assignment into groups (participation) that would be observed in experiments (Khandker et al. 2010).

In order to utilise PSM, there are two assumptions which need to be fulfilled: the conditional independence assumption (CIA) and what is known as common support. Conditional independence or “unconfoundedness” effectively means that the participation into either of the two groups (treated or non-treated) is determined purely by the observable characteristics and the outcomes are independent of the treatment status (Khandker et al, 2010). This is a very strong assumption is as it means that in order to use PSM, one must be able to observe all the characteristics that determine treatment; the exclusion of a control may also be likened to omitted variable bias. Khandker et al. (2010) highlight that if there are unobserved characteristics that determine whether or not a group is treated or not and the outcome under observation, then a “hidden bias” problem described by Rosenbaum (2002) may arise.

The second assumption is that there is “common support”: effectively for each covariate there is a positive probability of being in each group. For example if all households that were headed by someone who had no education experienced a health shock, then there would not be any cases in the non-shocked group. As a result, there would not be sufficient cases for comparison. Observations which fall out of the area of common support are dropped as only those within the area of common support can be used to undertake inference.

6.4.1 Implementation of PSM with KIDS

The use of PSM hinges on the satisfaction of the two assumptions of conditional independence and common support. The first step is to therefore consider the choice of covariates which will be used to calculate the propensity score. It should be noted at this juncture that there are differing views on what covariates should be included. For example, Caliendo and Keopeinig (2005) indicate that *both* outcome and participation covariates should be included in the calculation of the PSM, whereas Brookhart et al. (2006) suggest that covariates which are unrelated to the participation but related to the outcome should always be included. Austin et al. (2007) indicates that there are a number of different views amongst researchers using this technique: selection of covariates that only affect participation; only affect outcomes or as Caliendo and Keopeinig (2005) suggest covariates which are associated with participation and outcomes.

In *Chapter 4* the Mosely and Chen (1984) framework was used to identify the proximate determinants of health. These proximate determinants are factors which directly influence the probability of morbidity and mortality. Using this framework, one may consider maternal factors, environmental contaminations, nutrient deficiency, injury and personal illness control as a basis for constructing the propensity score.

The KIDS data includes information on the types of facilities a household has; such as water and sanitation sources. One would expect that household accessing water from a clean source such as a pipe would experience less morbidity relative to those who access water from sources which may have more exposure to contamination such as a well. Moreover, the type of toilet facilities a household has is also very likely to influence the spread of disease. Households with a flushing toilet are able to dispose of waste quickly relative to households who may use a pit or latrine and may therefore be expected to have a smaller risk of exposure to potential disease. The type of home someone may live in could also affect the ease with which people can protect themselves from the elements. People living in brick houses may find that their home is more watertight than someone living in a home made from wattle (where twigs and branches are laced together).

Indicators of household composition are also considered. If morbidity is disease related, larger households may be more likely to experience a serious injury/illness related shock as people living within the same area may transmit disease to other family members. The number of children and adults living within a household may also be relevant as children (particularly young children) face different risk factors to disease relative to adults as their immune system develops.

Indicators of health facilities and programmes should also be included as they are likely to influence the probability of a household experiencing health shocks. The available indicators include; distance and number of health facilities and whether or not there has been an immunization campaign within the local community since the last election. Two other potential variables for inclusion come from the community survey that complements the household survey; specifically whether the community itself has experienced a series of shocks since the last election. The two specific shocks that have been included as potential dummy variables relevant for the calculation of the propensity score are; human epidemic[238] and HIV. These indicators are considered for inclusion as communities which have experienced outbreaks and high HIV prevalence may be much more likely to include shocked households, ceteris paribus. Where no such epidemic was reported, the remaining communities have been coded as zeros accordingly.

The calculated propensity score is sensitive to the choice of covariates and will change as different variables are included or excluded and this may change the reported results quite substantially (Smith and Todd, 2005). If a variable that determines treatment is excluded then the assumption of conditional independence will be violated. One concern is based on the findings of Table 6.2.8 which indicate that households that have been shocked (illshock specifically) in 1998 are more likely to experience the same shock in 2004. The KIDS data enable this to be included in 2004 when calculating the propensity score but not in 1998 as the shock module did not exist in 1993. Therefore, when calculating the PSM for 1998, there is no way to include a control for being previously shocked. This particular violation is not directly testable (Khandker et al, 2010).

[238] Community leaders were asked to specify the type of human epidemic. The given responses were; conjunctivitis, diarrhoea, cholera and tuberculosis. In all cases, the epidemics were classified (by the community leader) as severe to very severe and estimated to affect at least half of the community in question.

6.4.2 Choice of matching

There are several different ways that the propensity score from the two groups, participants and non-participants, can be matched. Nearest neighbour (NN) matching is a technique which matches the propensity score of a treated case with the closest matching propensity score of a case from the non-treatment group. There are some additional considerations that may be taken when selecting such an approach; whether to match with or without replacement. Matching with replacement enables a case from the non-treatment group to be used multiple times (Khandker et al. 2010). Without replacement means that the untreated case may only be used once. Caliendo and Keopeinig (2005) indicate that using replacement, the quality of the matching increases and the bias decreases. They suggest that such an approach is very useful in cases where the distribution of the propensity scores is very different across the two groups. In such cases having suitable cases from the untreated group can be challenging, particularly as the observations become more limited once they are matched. Allowing for replacement ensures that the maximum possible number of potential cases for matching is available. However, the cost of allowing for replacement is that the estimated results may be sensitive to the sample of the untreated cases that may be used several times (Smith and Todd, 2005). A final issue to be aware of when using nearest neighbour matching is that the ordering of the cases may influence the estimation results. Therefore, the order of the cases should be randomised. Hence, the Stata command rsort[239] has been used for the purpose of the nearest neighbour estimations presented within this chapter.

An alternative matching method that may be used is caliper or radius matching. One of the difficulties that can be encountered with nearest neighbour (NN) matching is that even the closest neighbour may not be a very good match and the two scores could be quite different from one another (Caliendo and Keopeinig, 2005; Khandker et al. 2010). Caliper matching avoids this by setting a maximum distance between the two cases which is deemed to be acceptable which is thought of as a tolerance or threshold (Caliendo and Keopeinig, 2005; Khandker et al. 2010). The closest possible match effectively the nearest neighbour is then selected within the set tolerance level. According to Caliendo and Keopeinig (2005) the advantage of this method is that only cases that are within the set tolerance can be matched to one another, which should give

[239] Using a seed so that the results are reproducible.

rise to a better set of matches. A disadvantage of the approach is that if the scores are quite different to one another, then the number of suitable matches will fall. A variation on caliper matching is radius matching. This method uses all of the potential matches within the tolerance level rather than the closest match within the tolerance level. The benefit of the radius matching approach is that it uses more matches for comparison but avoids the issue of a possible bad match which could arise when using just the caliper approach (Caliendo and Keopeinig, 2005).

The next possible matching approach to consider is what is known as stratification or interval matching. This approach partitions the region of common support into smaller sections and compares the mean difference in outcomes for the two groups given a mean impact effect for each of the sections. The choice of how many partitions to have is set as five by default within Stata but can be changed depending on whether each partition is balanced, which can be tested using a post estimation test. If the test result suggests that the scores within the two groups in the interval are not balanced, additional intervals can be added until balance is achieved (Caliendo and Keopeinig, 2005).

Kernel and local linear matching is an alternative approach that differs from the three previous approaches in that it uses a weighted average of all the non-participants rather than a subset of non-participants to create the counterfactual outcome (Khandker et al. 2010; Caliendo and Keopeinig; 2005). The advantage of this technique is that the variance tends to be lower as more information (more non-participant cases) has been utilised to create the counterfactual case. However, just as in the case of nearest neighbour, potentially bad matches could still be included and it is therefore important to consider the area of common support when using this method (Caliendo and Keopeinig, 2005).

A final approach that can be considered is difference-in-difference matching and requires panel data or repeated cross sectional data which is stable over time (Khandker et al., 2010). This method does allow for unobserved differences between groups but assumes that such unobserved differences are time invariant (Khandker et al., 2010). By comparing the two groups before and after an intervention or the treatment, the difference in outcomes then estimates the effect of the treatment.

As there are several different matching methods to choose from, which in large samples should provide the same results according to Smith (2000), cited within Caliendo and Keopeinig (2005), one has a choice of which method to use with each method having a range of advantages and disadvantages. The choice is dependent on the data given, particularly in cases where there are only a few cases which can be used for matching according to Caliendo and Keopeinig (2005). Equally, they suggest if there are many potential cases which the treated can be matched with, in order to increase precision kernel matching may be a better choice. A recommended preliminary step by Caliendo and Keopeinig (2005) is to observe the area of common support as only cases within this region are used for the purpose of estimating the average treatment effect on the treated (ATT). The area of common support can be inspected by using a graph. In addition, propensity scores outwith the minima and maxima from the two groups can be excluded. Caliendo and Keopeinig (2005) note that inspecting the area of common support is more important when using kernel matching as all observations are used whereas with the nearest neighbour only the closest match is used and hence common support is generally satisfied.

6.4.3 Propensity score matching results

Table 6.4.1 provides a summary list of variables in addition to those presented in Table 6.2.1 which are specifically used for calculating propensity scores and related to a household experiencing a serious injury/illness shock. For presentation purposes as each indicator is a dummy variable no minimum and maximum columns are presented and nor is one for the median as it is zero in all cases with the exception of vaccines where it is 1. Table 6.4.2 provides some descriptive statistics on these additional variables for the 1998 data; 2004 descriptive statistics are presented separately as there are differences in the categories.

Table 6.4.1 Summary list of additional variables -1998 and 2004

Variable	Variable definition
Brick	dummy: equal to 1 if house is mainly made out of brick
Mudmix	dummy: equal to 1 if house is mainly made out of a mud and cement mix
Cement	dummy: equal to 1 if house is mainly made out of cement
Mud	dummy: equal to 1 if house is mainly made out of a mud
Wattle	dummy: equal to 1 if house is mainly made out of a wattle[240]
Other material	dummy: equal to 1 if house is mainly made out of a some other material which includes iron, tiles and cardboard.
Inside tap	dummy: equal to 1 if main source of drinking water is from an inside private tap
Outside tap	dummy: equal to 1 if main source of drinking water is from an outside private tap
Public tap	dummy: equal to 1 if main source of drinking water is from a public tap (this includes free and paid for sources)
Borehole	dummy: equal to 1 if main source of drinking water is from a borehole
River	dummy: equal to 1 if main source of drinking water is from a river
Other water source	dummy: equal to 1 if main source of drinking water is from some other water source which includes springs, wells and rainwater tanks
Flush	dummy: equal to 1 if the toilet in the household is a flush toilet
Pit	dummy: equal to 1 if the toilet in the household is some form of pit
No toilet	dummy: equal to 1 if there is no toilet in the household
Other toilet	dummy: equal to 1 if the toilet is some form of toilet such as a bucket or chemical toilet
Only 1998	
Vaccine	dummy: equal to 1 if the community where the household is located has experienced an immunisation campaign since the last election
HIV	dummy: equal to 1 if the community where the household is located has suffered from a HIV epidemic since the last election
Epidemic	dummy: equal to 1 if the community where the household is located has suffered from some kind of human epidemic since the last election
Dist clinic	Distance in kilometres to the nearest clinic (reported as zero if within community)
Dist hosp	Distance in kilometres to the nearest hospital (reported as zero if within community)
Dist pharm	Distance in kilometres to the nearest pharmacy (reported as zero if within community)
2004 dwelling	
Shack	dummy: equal to 1 if home is a shack
House	dummy: equal to 1 if home is a house
Hut	dummy: equal to 1 if home is a hut
Hostel/Outbuilding	dummy: equal to 1 if the home is either a hostel or outbuilding
Combination	dummy: equal to 1 if the home is a combination of buildings
Other home	dummy: equal to 1 if the home some other type of home

[240] Wattle consists of branches and twigs.

Table 6.4.2 Descriptive statistics 1998

Variable	All households		Shocked[241]		Non-shocked	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Brick	0.292	0.455	0.286	0.453	0.294	0.456
Mudmix	0.205	0.404	0.220	0.415	0.201	0.401
Cement	0.217	0.412	0.216	0.412	0.218	0.413
Mud	0.179	0.384	0.158	0.365	0.183	0.388
Wattle	0.086	0.280	0.108	0.311	0.080	0.272
Other material	0.021	0.145	0.012	0.111	0.024	0.152
Inside tap	0.290	0.454	0.212	0.409	0.309	0.462
Outside tap	0.137	0.344	0.183	0.387	0.126	0.331
Public tap	0.130	0.336	0.129	0.335	0.130	0.336
Borehole	0.098	0.298	0.154	0.361	0.084	0.278
River	0.231	0.422	0.249	0.433	0.227	0.418
Other water	0.115	0.319	0.075	0.263	0.125	0.330
Flush	0.344	0.475	0.301	0.462	0.353	0.478
Pit	0.529	0.499	0.577	0.495	0.517	0.500
No toilet	0.112	0.322	0.104	0.306	0.120	0.326
Other toilet	0.010	0.010	0.012	0.111	0.009	0.096
Vaccine	0.953	0.211	0.936	0.245	0.957	0.202
HIV	0.194	0.396	0.137	0.344	0.201	0.407
Epidemic	0.135	0.342	0.154	0.361	0.131	0.337
Dist clinic[242]	3.626	4.041	3.438	3.465	3.671	4.170
Dist hosp[243]	25.022	20.487	27.107	21.827	24.524	20.136
Dist pharm[244]	17.233	17.400	16.636	18.031	17.367	17.264

Source: KIDS (1998)

An examination of Table 6.4.2 reveals some slight differences between the two groups in terms of their household services and potential exposure to disease. Taking the material which the house is built from; it is observed that very similar proportions of shocked and non-shocked households live in homes made from brick and cement. A very slightly larger proportion of shocked households live in homes which are constructed of a mud and cement mixture, wattle or mud relative to non-shocked households.

A much more noticeable difference in household services is noted with regards to the source of drinking water households have. Twenty-one percent of shocked households have an inside tap as their main source of drinking water in comparison to thirty-one

[241] Shocked in this context is specifically related to “illshock”.
[242] Reported for 1152 cases out of 1212.
[243] Reported for 1023 cases out of 1212.
[244] Reported for 981 cases out of 1212.

percent of non-shocked households which is a substantial difference between the two groups. Moreover, larger proportions of shocked households source their water from rivers and boreholes relative to non-shocked households. A comparison on the types of sanitation households shows a similar pattern; a much larger proportion of non-shocked households have flushing toilets and fewer use pits relative to shocked households.

Table 6.4.3 Descriptive statistics 2004

Variable	All households		Shocked[245]		Non-shocked	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Shack	0.022	0.146	0.015	0.125	0.023	0.153
House	0.441	0.497	0.307	0.462	0.482	0.500
Hut	0.112	0.316	0.169	0.376	0.169	0.376
Hostel/outbuilding	0.014	0.120	0.016	0.125	0.014	0.119
Combination	0.392	0.488	0.466	0.500	0.366	0.482
Other	0.017	0.129	0.026	0.161	0.014	0.119
Inside tap	0.300	0.458	0.222	0.417	0.325	0.469
Outside tap	0.211	0.408	0.234	0.424	0.205	0.404
Public tap	0.198	0.399	0.233	0.424	0.189	0.392
Borehole	0.075	0.263	0.095	0.294	0.070	0.255
River	0.148	0.356	0.159	0.366	0.143	0.351
Other water	0.065	0.247	0.058	0.235	0.064	0.244
Flush	0.367	0.482	0.265	0.442	0.400	0.490
Pit	0.560	0.497	0.656	0.476	0.529	0.500
No toilet	0.074	0.261	0.079	0.271	0.072	0.258

Source: KIDS (2004)

Vaccination programmes are likely to be targeted at children, although the information within the KIDS community questionnaire does not specify the type of immunisation programme that has taken place within the community since the last election and therefore it is not possible to confirm any details on the programme/s. Table 6.4.2 indicates that overall the vast majority of communities have had some kind of immunisation programme since the previous election; with a slightly larger proportion of non-shocked households benefitting from an immunisation programme. A much larger proportion of non-shocked households have been affected by HIV relative to shocked households, which may suggest that the health shocks reported by the household are not HIV related.

[245] Shocked in this context is specifically related to "illshock".

The 2004 dataset does not include a community survey. Therefore there are no details on epidemics or vaccinations or health facilities generally within the community. The category of housing was also changed in 2004; rather than asking about the type of material the home was made from, respondents were asked to report the type of dwelling which is noted in Table 6.4.1.

Table 6.4.3 indicates that the type of home people live in does vary across the two groups, with a larger proportion of non-shocked households living in a house relative to non-shocked households. Another notable difference between the two groups is that a larger proportion of non-shocked households have an inside tap as their source of drinking water; it would appear that overall shocked households may have access to potentially more risky water sources. A similar pattern is found for the sanitation access households have; a larger proportion of shocked households use a pit as opposed to a flush toilet which is more common amongst non-shocked households.

Following the calculation of the propensity score, it is necessary to examine the matching to assess whether the two groups are balanced; the basis of the balancing test is to confirm that the conditional independence assumption is satisfied, which is a necessary requirement of PSM (Henrich et al. 2010). The idea is that after matching there are no additional controls that could be added that would improve the estimation and there should be no statistically significant differences between the means for the controls across the two groups after matching (Caliendo and Keopeinig (2005). If after completing the match, there is evidence that the groups are unbalanced then there is a need to reconsider the participation model and or use an alternative matching algorithm (Henrich et al. 2010; Caliendo and Keopeinig, 2005). Unfortunately it is also possible that being unable to achieve balance between the two groups may suggest that the conditional independence assumption is violated according to Smith and Todd (2005).

The initial participation model used for the calculation of the propensity score includes the following controls; household size, gender of the household head and an indicator for whether they had any form of education and indicators for their employment, the type of material used to construct the house, the water and sanitation source, whether there had been any incidence of HIV within the community, whether there had been a vaccination programme within the community and indicators for distance one would travel to various medical facilities (hospital, clinic and pharmacy) in the case of 1998. Table 6.4.4, column (1) uses nearest neighbour matching, with one neighbour, without trimming and with replacement. Replacement has been included as Caliendo and Keopeinig (2005) suggest that such an approach provides a better quality of matching. The ATT from this initial estimation is reported in Table 6.4.4 column (1) and Figure 6.4.1 illustrates the area of common support between the two groups. However, before proceeding it is important to consider some balancing tests.

Overall the variable t-test results on the matched means for the majority of the variables are statistically insignificant after matching in all cases which is encouraging. However, Caliendo and Kopeinig (2005) suggest that after matching the reported standardised bias on the observables should be less than 5 percent; this is not the case for a number of variables; the dummies representing the building material of the house, the source of drinking water and sanitation facilities and the indicators on epidemics, vaccination programmes and the distance to a pharmacy and clinic. An examination of the distribution of the average absolute bias before and after matching indicates that before matching, the mean bias was 8.7% and after matching this falls to 5.0%. Rubin's B and R may also indicate whether there is balance between the two groups; the reported Rubin's B in this case is 48.2 which is in excess of the recommended maximum of 25 by Rubin (2001) but the Rubin's R is 0.89 which is within the recommended boundary of 0.5-2[246]. Therefore, there is evidence to suggest that the two groups are not well as balanced as one would like. A common suggestion following balancing test failures is to include the higher order terms (such as the square) of a variable or an interaction (Lee, 2013; Caliendo and Kopeinig; 2005; Dehejia and Wahba, 1999) as well as considering alternative matching algorithms. Common support is evidenced as shown in Figure 6.4.1.

Therefore, the next consideration I make is to increase the breadth of nearest neighbour matches from one to five to see if this may improve the balance of the groups. Five is selected as it suggested by Khandker et al. (2010) as a suitable number of neighbours. Increasing the number of neighbours to five, indicates that the groups are much more

[246] Stata help file on pstest

closely balanced overall but there is still evidence that balance is not achieved as the mean bias on a number of coefficients remains above 5 percent following matching (on sanitation facilities as before and also on community indicators such as vaccination programmes and epidemics). Rubin's B (24.3) and R (0.77) are both within the recommended limits following the use of five neighbours. The ATT from this specification is reported in Table 6.4.4 column (2). The improvement in balance results in a lower estimated ATT of -0.301 and again is reported as being statistically different from zero but as was the case previously it has a very wide confidence interval indicating that the log percentage ATT could be between [-0.603, -0.001] which is very large indeed and therefore very imprecise.

Table 6.4.4 Propensity score – 1998

	(1)	(2)	(3)	(4)
ATT	-0.438 (0.180)	-0.301 (0.151)	-0.238 (0.129)	-0.237 (0.135)
N untreated (on support)	629	629	629	637
N treated (on support)	161	161	161	160
Total N[247]	790	790	790	797

Source: KIDS. Standard errors in parentheses[248]

One of the issues with nearest neighbour matching is the possibility that the closest match(es) may not be very good. Therefore a caliper can be used to try and improve the quality of the match. The specific choice of caliper (which is the tolerance level) is subject to debate (Caliendo and Kopeinig, 2005) as it is not known what level of tolerance is appropriate a priori. A caliper of 0.05 is initially considered but the results indicate that imbalance remains. The use of different width calipers does not improve the problem of imbalance.[249] A final consideration presented uses a kernel match: specifically an Epanechnikov kernel, with a bootstrap and 1000 replications; overall this specification balances the two groups. The results from the kernel specification are reported in Table 6.4.3 column (3). The estimated ATT is smaller in this case, with a (log) percentage difference of -23.8%. This translates to 21.2 percent in standard

[247] This is the total number of cases, including those off support.
[248] Abadie and Imbens (2006) in the case of nearest neighbour and bootstrapped to 1000 replications in the kernel estimation.
[249] Caliper match results are not reported.

percentage terms. However, the confidence interval for this is very wide and indicates that the ATT could be very large or relatively small.

In the nearest neighbour specifications presented in Table 6.4.3 column (1) and (2) the reported t-statistic indicates that the percentage difference in (log) earned income is statistically different from zero. However, the results also indicate that the ATT is very imprecisely measured and there is imbalance between the two groups. Therefore, whilst there is evidence that health shocks have a negative effect on household income; it is not possible to indicate the magnitude of this with a degree of precision, especially given that there are issues associated with the balance between groups.

Given that the kernel specification achieves better balance, I consider a variation on the measure of education; using a series of levels as opposed to a simple dummy to see how this may affect the results. The estimated ATT is very similar as reported in Table 6.4.4 column (2). I also consider different indicators for medical facilities; rather than distance, the number of facilities within the community. The results are broadly similar and therefore are not presented.

Figure 6.4.1 Visual inspection of common support using specification in column (3) Table 6.4.4

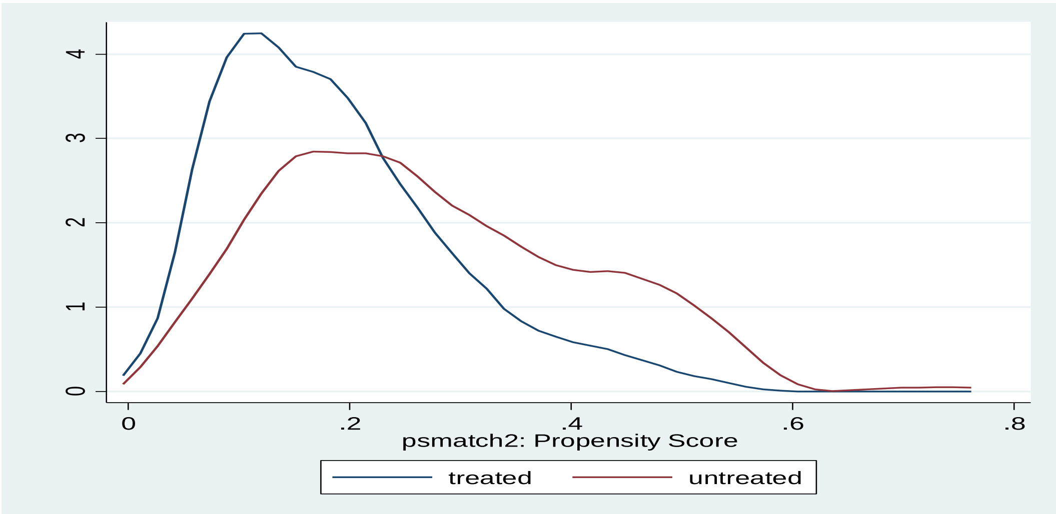
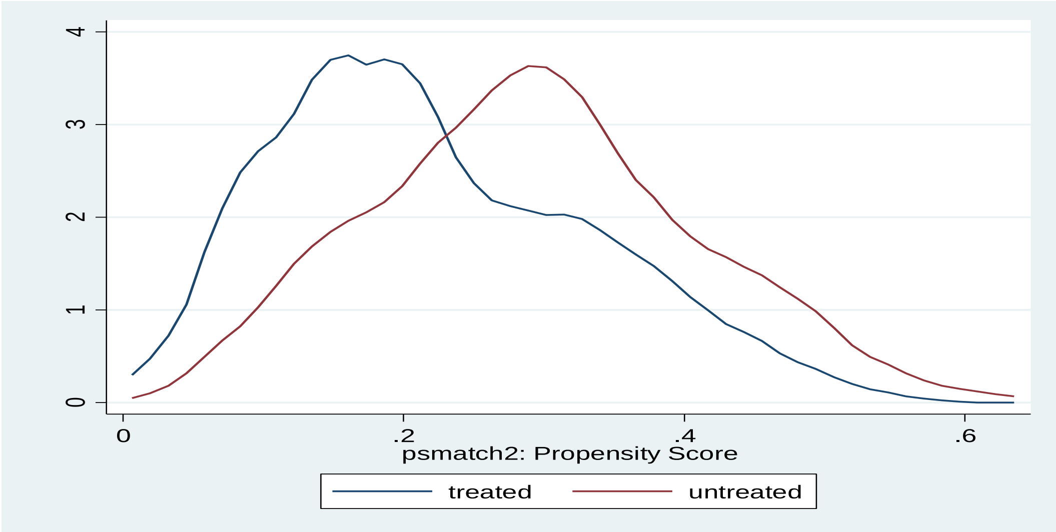


Figure 6.4.2 Visual inspection of common support using specification in column (3) Table 6.4.5



The discussion will now move on to an examination of the 2004 data to see if better estimates of the ATT may be obtained from the inclusion of an indicator for previous shock status. As before, a review of the balance between the two groups will take place before moving on to an examination of the ATT. The controls included within the participation model are broadly similar to those used for the 1998. The exception is the inclusion of whether the household experienced a shock in 1998 and the exclusion of any community characteristics as such data was not available in 2004.

The results from the participation model indicate that experiencing a shock (illshock) in 1998, is associated with the probability of experiencing a shock of the same type in 2004 and this effect is statistically significant at the 5 percent significance level. In the 1998 model it was not possible to include this indicator as no such data was available. What it does highlight is that households who are shocked are ceteris paribus more likely to be shocked again. Experiencing a health shock in 1998 raises the probability of experiencing the same shock in 2004 by 11.07[250] percentage points; this result is statistically different from zero and is reported with a 95 percent confidence interval of [0.025, 0.196]. Using one nearest neighbour, with replacement and no trimming, the first ATT calculated is 0.381 and reported in Table 6.4.5 column (1). This suggests that the treatment effect (experiencing a health shock) is positive. However, the result is not statistically different from zero. Moreover, there is evidence of imbalance on a number

[250] Calculated using the mfx command at the mean.

of variables when examining the post-matching mean bias on the controls for the type of home and water source someone may have as well as whether or not the household head has any education. The t-statistics on the difference in means in all cases are statistically insignificant but Rubin's B (42.5) is outwith the permitted range post matching.

An alternative NN match is then considered in Table 6.4.5 column (2), expanding the number of neighbours to five to see whether this may lead to an improvement in the balance of the two groups. Increasing the number of neighbours used for matching improves the overall balance of the two groups in some areas (reducing the mean bias on some coefficients post matching) but the reported ATT remains positive, although again this result is still not statistically different from zero. Rubin's B also remains larger than 25 at 28.6.

Using the same model but with a caliper of different widths makes no difference to the results and therefore the ATT from this matching algorithm is not reported. A final consideration that was used with the 1998 data was that of a kernel, specifically the Epanechnikov kernel. This matching method ensures that there is overall balance within the two groups. The reported t-statistics and Rubin's B in this case are indicative of balance. However, there is again no evidence to support the view that the ATT is not statistically different from zero and the 95 percent bias corrected confidence interval is very wide [-0.142, 0.362][251] and imprecise. Due to the imprecision of the results, I am unable to conclude that there is evidence to support the view that the ATT is even negative in 2004.

The results in specifications (1)-(2) of Table 6.4.5 suggest that there is imbalance between the two groups. Therefore, a consideration is to select different controls for the participation model to see if this can address the problem. Education of the household head was included as a simple indicator in the results presented in Table 6.4.5 for (1) and (2). Therefore, the propensity score and the ATT were recalculated using differing levels of education to observe whether this improved the balance within the groups. This change to the participation model does not improve the balance in the NN matches.

[251] Bias corrected 95% confidence interval.

The original kernel match using the simple indicator of education is replaced and the result is reported in column (4) of Table 6.4.5. The estimated ATT remains positive and again is not statistically different from zero

Table 6.4.5 Propensity score – 2004

	(1)	(2)	(3)	(4)
ATT	0.381 (0.245)	0.224 (0.166)	0.128 (0.154)	0.112 (0.135)
N untreated (on support)	462	462	462	462
N treated (on support)	143	143	145	142
Total N[252]	607	607	607	607

Source: KIDS (2004). Standard errors in parentheses[253]

Overall the PSM results provide some very limited evidence that there are differences in the (log) earned income of households in the KwaZulu-Natal province in 1998 and that shocked households have lower earned incomes. However, the magnitude of this is very imprecisely estimated. The NN matches suffer from balancing problems; although there is a lack of consensus on the value of such tests and how much attention should be paid to them (Lee, 2013), the kernel matching algorithm is able to achieve balance. The 2004 results however are unable to indicate that the treatment effect is even negative.

6.5 Directly-reported income losses

As discussed in the introduction to this chapter, one of the unusual features of the KIDS dataset is that households report an estimate of their monthly income lost as a result of the associated health shock. In principle, the benefit of using a direct estimate, reported by each household, is that I do not need to account for additional control variables. Implicitly as each household reports their loss, individual characteristics are controlled for in each case; one is not reliant on only observable characteristics. There are a couple of issues that need to be considered. The first is that respondents were asked to estimate the monthly loss of income associated with a shock. To keep things simple in the example that follows, it will be assumed that only one shock has occurred. In cases where income is regular, such as a monthly salary, if an individual subsequently experienced a health shock (assuming no sick pay benefits), it would be clear what the

[252] This is the total number of cases, including those off support.
[253] Abadie and Imbens (2006) in the case of nearest neighbour and bootstrapped to 1000 replications in the kernel estimation.

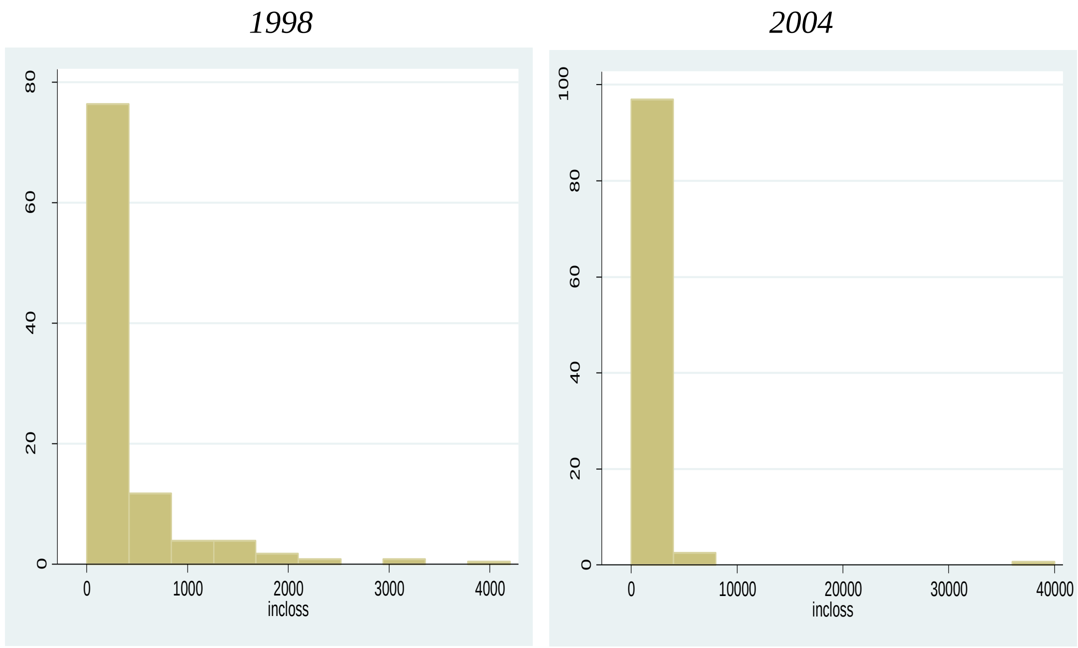
monthly income loss would be; the difference between what was received before and after the shock. The most recent reference period in such a case would be the salary payment in the previous month. Even with this very simple scenario, however, the estimation becomes more complicated if salary payment amounts were less regular; one may not be able to readily distinguish between changes due to shocks and changes due to other factors unrelated to shocks such as; working more hours, bonuses, different payment periods and so forth.

The next complication is that households *may* report a loss which is not particularly associated with the specific shock they were asked about. If households effectively aggregate income losses, across multiple different types of shocks, then there could again be considerable measurement and thus reporting error. As an example, one of the possible shocks experienced by a household is abandonment/divorce. If this occurred during a similar time period to the health shock, it may be questionable as to whether households can separate out the two scenarios; attributing loss of income due to a health shock independently of any losses which may be associated with the abandonment/divorce. In order to assess whether this could have occurred, the reported losses have been inspected for households reporting one or more shocks. It does not appear that households have reported the same losses across different shocks. This may therefore be indicative that households were in fact able to distinguish shocks from one another.

A final consideration is related to the validity of the data on reported losses, particularly where reported losses are very large. Given that households are very unlikely to know what their counterfactual income would have been, the estimate of losses could be inflated or under-estimated. An examination of the distribution of losses may provide some indication as to any potential outliers and the general pattern of losses experienced. Figure 6.5.1 indicates that around three quarters of households experiencing a health shock[254] indicate losses which are under 400 Rand in 1998 and 600 Rand in 2004. The very large losses appear to be outliers as the overall distributions indicate much more modest losses.

[254] Serious injury/illness related shock.

Figure 6.5.1 Distribution of income losses



Source: KIDS (1998) and KIDS (2004)

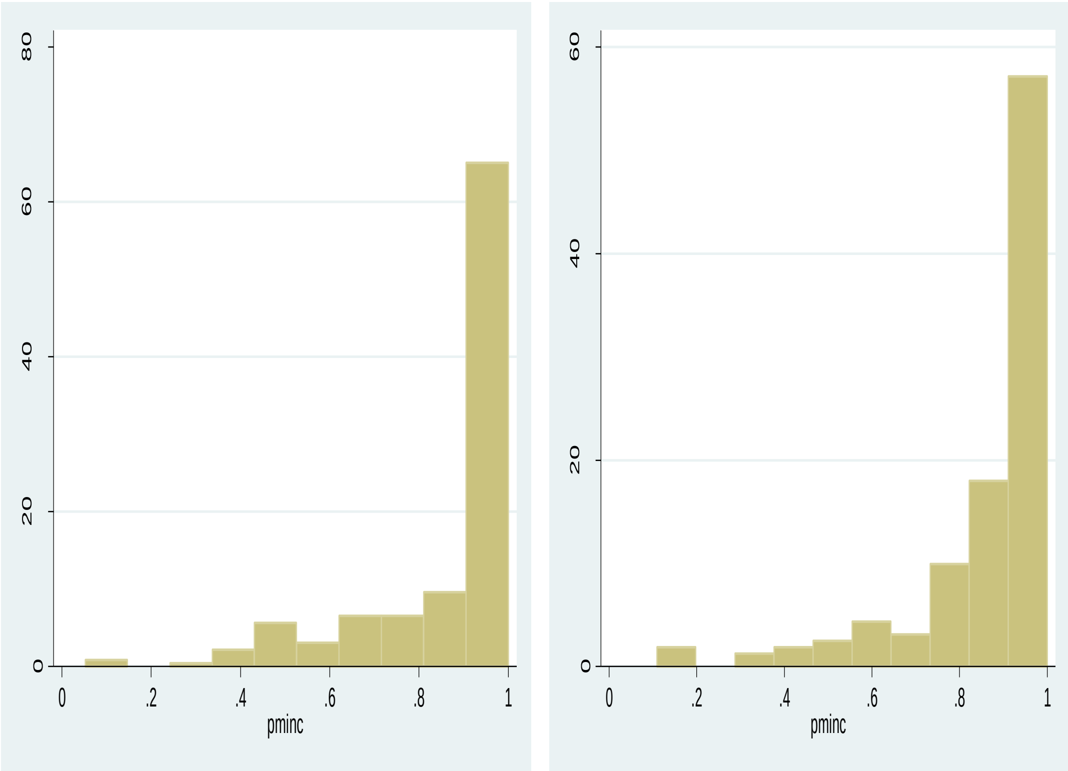
To examine these differences in income more closely, a measure of post illness monthly income has been developed; what household income would have been in the absence of a health shock which effectively utilises the value of total income that they have reported at the time of the survey. Total income is being used rather than earned or unearned income as the question related to the losses does not specifically indicate whether respondents should differentiate the source of the loss.

Post illness monthly income can be defined as the income (T) by the household relative to what their income would have been in the absence of a shock (hypothetical income). Households not experiencing a shock (illness) will receive 100 percent of their hypothetical income as they do not experience any losses. Households who have been shocked and experience losses (L), will receive a proportion (Pminc[255]) of their hypothetical income where; Pminc = T / (T + L).

[255] Where Pminc is the proportional post illness monthly (earned) income.

Figure 6.5.2 Proportionate post illness monthly income relative to hypothetical income

1998 *2004*



Source: KIDS (1998) and KIDS (2004)

Using this approach, it is observed using Table 6.5.2 that in 1998, a quarter of households who experienced a shock of "serious injury or illness" lost at least one-fifth of their monthly income whilst 50% or more of households lost nothing at all (represented by the spike at the right of Figure 6.5.1 in the left hand panel).

Comparing the results from 1998 with those of 2004, it is observed that the emerging picture looks broadly similar. In 2004, half the households lost 7% or more of their monthly income and twenty-five percent lost at least one-fifth of their hypothetical monthly income. These results indicate that for some households losses associated with a serious injury/illness shock can be relatively large, whilst the majority are largely unaffected.

Table 6.5.1 Proportionate post illness monthly income relative to hypothetical income

	1998	*2004*
Percentiles		
25%	0.810	0.823
50%	1	0.938
75%	1	1
Mean	0.876	0.869

Source: KIDS

In both 1998 and 2004, catastrophic losses (loss of over half the normal monthly income) are relatively rare; around 7 percent of cases in 1998 and 6 percent in 2004. A direct comparison of the duration of a serious injury/illness shock in 1998 and 2004

6.5.1 Distributional issues and transitions over time

Figures 6.5.1 and 6.5.2 show that the distributions of income losses are highly skewed. The majority of households are largely unaffected by health shocks; but in a small number of cases, income losses can be large. An examination of all households' transitions between the two survey periods indicates that the majority of households do not change their relative income[256] position between 1998 and 2004. Seventy percent of households who were below the median income in 1998 remain below the median income level in 2004. Only 27 percent of households who were initially below the median income level increased their income to above the median in 2004. Similarly, of the households that were above the median income in 1998, twenty-nine percent now find that they are in a below median income position by 2004. The vast majority of households who were above median income in 1998 (73 percent) remain so in 2004. A closer examination of the households who experienced a shock[257] reveals that changes in the relative median income position of households do not differ substantially between shocked and non-shocked households.

Clearly, in the analysis just presented, a key issue is the potential reporting error of both the total income figure (from earned and unearned sources) and the reported income loss associated with the shock. The under or over-reporting of the two figures could lead to different results. Moreover, in terms of the relative income position of households in

[256] Based on total monthly income which is inclusive of income from both earned and unearned sources.
[257] The shock is defined here as a household which has indicated that they experienced a "serious injury or illness".

1998 and 2004, this figure is potentially very sensitive to when the household was surveyed as previously mentioned. Households who have had a particularly poor month due to a poor harvest or an unexpected cost if they have a business may have a reported total income that month which is much lower than what they would report in more "normal" circumstances. Likewise, a household that has perhaps received a large windfall would report an income which would be much higher than normal. Therefore, as previously suggested, a more appropriate measure might be the income received by the household over a slightly longer period as such fluctuations may be smoothed out, giving a more realistic picture of household incomes.

Table 6.5.2 Transitions in median income position between 1998 and 2004

	Median income 2004	
Median income 1998	% below median	% above median
% below median income	70.47	29.53
% above median income	27.25	72.75

Source: KIDS (1998) and KIDS (2004)

6.6 IV estimation

As in *Chapter 5,* one must consider the possibility that health shocks are endogenous. In such cases an alternative approach is to use IV estimation, assuming suitable instruments can be identified. Within the KIDS data there are some possible instruments that may be considered, which may be considered to be correlated with the incidence of a health shock and independent of the error term in the income equation. In this section I briefly discuss the results of using IV methods with the KIDS data; detailed discussion is omitted because of space constraints and because of the inconclusive results and in particular the problems with weak identification.

The community survey for KIDS presents some potential variables that may be considered to instrument for health shock; these data are reported only for 1998 and not for 2004. The potential instruments that are initially considered are the distance in kilometres to the nearest public hospital, pharmacy or clinic, and the number of health facilities[258] in the area.

[258] Public hospital, pharmacy and clinic.

In the event of a health shock, one may seek to receive medical attention and the closer such facilities are to households, potentially the easier it may be for them to access such facilities and in turn be treated and thus avoid or reduce the severity of the shock. The distance one may have to travel to the various types of health facility or the numbers of health facilities within an area do not include any indication of affordability. In both cases, the potential validity of the instrument rests on whether the instruments are sufficiently correlated with the health shock but uncorrelated with the error term.

Whilst it is very possible that health facilities may not have changed between the two survey periods (1998-2004) there is no way of knowing this with any certainty and as such, IV estimation can only potentially be undertaken with the 1998 data.

Partly for reasons of space and partly because the results of specification tests were unsatisfactory, I do not report the IV estimation results in detail. The specification tests used were the Hansen J test for the exogeneity of the instruments, and a first-stage F statistic to test for whether the instruments were weak or strong (i.e. the strength of the correlation with the health shock variable).

The results of the Hansen J test were mixed. In some specifications, the null hypothesis of exogenous instruments was not rejected, and in other specifications it was. However, the main problem was that of extremely weak instruments. For example, when using the distance to health care facilities as instruments, the first stage F-statistic was very small in the case of a serious injury/illness shock, with a reported F-statistic of 1.085, and 3.106 in the case of a deathshock.

From a practical point of view the KIDS dataset (much like the IHS2) has a limited set of variables that may be considered as potential instruments. The number of health facilities within a community *potentially* enables people to have access to health care professionals and medical inputs which can be used to maintain an individual's stock of health and may therefore act as an instrument for health shocks. The problem with this proposal is that just because there are health facilities, it doesn't mean that people use them (as in the case of distance to health facilities). If this were the case then there would be a weak correlation between health status and the number of facilities. Another problem with the number of health facilities within an area is that it could potentially be endogenous: there could be more facilities if the health of people in the area is very poor which led to greater state investment to provide such facilities for people.

I experimented with using number of health care facilities as an additional instrument in the IV estimations. Unfortunately, this alternative instrument is also quite weak. As in the previous case, the reported F-statistics from the first stage results are again very small, 1.107 in the case of a serious injury/illness shock and less than 1 in the case of a death shock. Moreover, the reported coefficients on each of the health shocks in the second stage results are reported with very large standard errors resulting in a lack of clarity on the estimated effect of health shocks on household income.

6.7 Conclusion

The direct estimates of income losses associated with a negative health shock reported in section 6.5 show that income losses may not be trivial for a small proportion of households but on the whole households do not appear to experience declines in income over the longer term. Using PSM as an alternative way of controlling for variation across the sample indicates some limited evidence to support the view that health shocks negatively affect household income. However, the results are not compelling overall, particularly in the case of 2004.

Ideally, income and shock data for the past year would have been preferable to investigate whether health shocks affect income in the short run. From the point of view of recall, more recent data would certainly be expected to be more reliable with less error. The reported income is also potentially highly sensitive to when the household was surveyed as only monthly data has been collected, yet there may be a high degree of fluctuation in their income over time which cannot be accounted for. A more appropriate examination of shocks and income would be to consider longitudinal data that can be broken down into shorter time periods to observe the possible evolution of shocks over time.

Households may also not accurately report shocks; it could be as simple as poor recall or as the injury/illness shock is subjective and associated with the perception of the respondent of particular events; what one may consider to be serious may not be viewed in the same way by another. This could lead to over reporting in some cases and under reporting in others. One of the biggest constraints of the economic shocks module is that multiple shocks of the same type cannot be reported. Ideally, a control for this could be desirable. Particularly given that there is evidence to suggest that shocked[259] households are more likely to experience a shock again in future.

An additional limitation to the KIDS data is that shocks are not reported by individual; ideally controls for the type of individual would also be preferable. The literature exploring the effect of shocks on economic outcomes clearly indicates that the results are sensitive to who is directly affected by the shock and this is a limitation of the KIDS data.

A particular difficulty encountered with the 2004 data is also the lack of identification of the household head which is subsequently used for analysis. This results in a much smaller sample in the case of 2004 as the household head cannot be readily identified. The selection of an appropriate alternative; oldest male; oldest female or oldest working individual could be considered in future work, but the data constraints on the key variables; shocks and income would remain.

In conclusion, a range of data at both the household and individual level over time is necessary to construct economic models that can accurately evaluate the impact of different health outcomes on well-being.

[259] In the case of injury/illness based shocks.

Chapter 7: Conclusion

This thesis seeks to address three aims: to examine the determinants of child mortality in Malawi using the IHS2 dataset, to consider the extent to which morbidity in adults affects wage rates in Malawi and finally to examine the extent to which health shocks may affect household income using the KIDS dataset.

Chapter 4 examines whether household income influences the incidence of under-five and infant mortality using the Mosley and Chen (1984) framework. The empirical work undertaken in *Chapter 4* utilises the IHS2 dataset in contrast to previous work undertaken by Manda (1999). The proximate determinants of child mortality are well established within the literature, but differences in the relative role of these determinants exist across countries and in particular there are different views on the role of income.

The empirical work presented in *Chapter 4* finds no evidence to support the view that increasing household income would reduce the probability and incidence of experiencing an under-five or infant mortality in Malawi. This result is robust across a range of different specifications including the use of different measures of household income. In addition to the inclusion of income, an additional indicator is considered within the model; the use of bed nets. Given that Malawi has a high incidence of malaria, the inclusion of an indicator for this preventative measure is important. The use of bed nets reduces the probability and incidence of morality across all specifications and the effect is relatively large given the overall rate of mortality.

The relative role of education is also pertinent across all specifications and demonstrates that possessing at least some education will reduce the rate of child and infant mortality. Education levels within Malawi overall are relatively low; therefore the continued expansion of education for young women as part of the wider development strategy is important. Education is frequently highlighted within the literature as being an important determinant of child health outcomes. There is a consensus that many determinants of health operate through education; such as the decisions made by mothers, although the nature of this relationship is unclear and represents an area of future research.

An extension to the existing literature is a consideration of the potential differences in the determinants of child and infant mortality given the cause of death. The results presented in Table 4.4.5 highlight that possessing at least some education can reduce the probability of experiencing an illness related death. In contrast there is no such evidence to support the view that education reduces non-illness related deaths. I propose that non-illness related deaths may be outwith the control of the mother and therefore determined by others. In particular there is evidence that pre-natal care (as measured by whether the mother attended a clinic during pregnancy) is a key factor in reducing the incidence of non-illness related deaths. Ensuring women have access to professional advice during their pregnancies enables healthcare professionals to monitor and support women who may face challenges at birth. Birth complications are one of the primary factors for infant death so identifying vulnerable women in advance may help to prevent this type of death.

One of the key challenges of using the IHS2 rather than the DHS to explore child mortality in Malawi is the lack of data on some key indicators that would be desirable. Notably, information on the rate of breastfeeding, birth order and information about each birth rather than only the most recent birth would be desirable. In addition, the IHS2 only identifies deaths that have occurred within the last two years and consequently restricts the sample size for analysis. A challenge that is experienced across all areas of the thesis is related to the indicator of income used. Income is reported by households over a relatively short period of time within the IHS2 and may therefore be sensitive to when the household was surveyed. A more preferable indicator would be to have income data over time so that fluctuations in income could be taken into consideration.

The aim of *Chapter 5* was to examine the extent to which morbidity may affect wages in Malawi. Morbidity represents a decline in the stock of health capital which is a part of an individual's human capital. Health is measured within this chapter using a self-reported indicator referred to as "disabled days" which considers the incidence and duration of morbidity within a two-week reference period. The results presented within *Chapter 5* indicate that male workers who have experienced recent morbidity receive a wage that is nine percent lower than an equivalent healthy male, ceteris paribus. Taking into consideration the time one may experience morbidity for, rather than just the incidence, it is found that the daily rate falls by one percent per disabled day. The sample for female workers is smaller and across the specifications presented there is no evidence that morbidity negatively impacts the daily wage rate received.

Returns to education at all levels are large and indicate that investment in education is important for Malawians. As previously noted, the level of education within the country is relatively low. Therefore expansion of education at the primary level in the first instance would be a priority and fits in well with other development policy objectives such as reducing the rate of child mortality within the country.

In addition to the current literature that explores the relationship between health and wage rates, the thesis also considers the effect of morbidity on the wage rate received by casual labourers which is known as ganyu labour in Malawi. I propose that as ganyu labourers can move from place to place, in workplaces that may be isolated from one another, individuals may be able to disguise their general health status and receive the same daily rate as their equivalent healthy counterpart. The results presented in Table 5.3.7 indicate that one cannot rule out the possibility that morbidity has no effect on ganyu rates.

One of the limitations of the analysis presented within *Chapter 5* is related to the problem of dual causality running between income and health outcomes. In order to deal with the potential endogeneity of the health measure used within regression work, IV estimation can be utilised. Crucially however, IV estimation requires that the instruments selected are sufficiently correlated with the endogenous variable. The IV estimation presented considers the distance one must travel to a clinic and doctor as potential instruments for disabled days, as guided by the work of Shultz and Tansel (1997). The IV estimation suffers from various specification problems including weak instruments, making any IV estimation using these instruments unreliable. In an attempt to address this problem, alternative instruments are explored, but these were also found to be too weak for reliable IV inference.

In addition to the modelling limitations, the indicators used within the analysis also have some limitations. Wages are only observed for those that have been recently working in the labour market. Individuals who have withdrawn from the labour market due to severe morbidity are not captured. Therefore the estimated effect of morbidity only considers those within the current working sample and may ignore those who have been most severely affected by morbidity. In future research and survey design, a question which seeks to elicit information about labour market participation over a longer period could be particularly useful.

The self-reported measure of health and wages used for the analysis undertaken also rely on recall and may be subject to reporting error. A relatively short reference period for morbidity should in principle reduce this problem. However, the incidence of morbidity may also be sensitive to when the individual was surveyed, given the nature of malaria within Malawi. The most recently reported income measure may also be subject to seasonal fluctuations that cannot be controlled for given the data available within the IHS2. In order to undertake future research in this area, one needs to be able to take into account seasonal fluctuations that may be relevant.

As in the case of *Chapter 4*, malaria is found to be a leading cause of morbidity in Malawi. Therefore the policy focus should look to increasing the provision of prevention of the disease in both adults and children. Ensuring that people have access to bed nets and affordable antimalarial drugs is essential. The prevention of morbidity enables people to maintain their stock of health capital and maximise the amount of working time available. In so doing one may be able invest in health inputs such as medical care to increase the stock of future health capital which has benefits for individual and household welfare. Future research needs to consider survey design and the availability of indicators that may be used to instrument for health, given the problem of endogeneity. Knowing that health is endogenously determined is one thing; being able to instrument for it using survey data is another matter. Although in practice the issue of identifying suitable instruments is not uncommon, it is a limitation to this type of research.

The aim of the final empirical chapter of this thesis was to examine whether health shocks have a negative impact on household income. The model framework utilised within this chapter is the Grossman (1972) model which illustrates that health is a capital good which can be used to produce healthy days. The incidence of health shocks reduces the stock of health capital, resulting in a loss of time for productive activities.

The results from the PSM model used for the 1998 data indicate that illness/injury related health shocks do have a small impact on current household income on average. For the 2004 data, I cannot exclude the possibility that health shocks have no impact on household income. A novel contribution to the literature is that as well as using models to estimate the effect of health shocks on household income, directly reported income losses can also be explored. An examination of the directly reported income losses associated with an injury/illness related shock indicate that in the majority of cases, income losses are small in both the 1998 and 2004 data. However, there are a small number of cases that indicate that their income losses as a result of experiencing a health shock have been much more substantial.

A limitation to the analysis presented within *Chapter 6* is the lack of information available to differentiate health shocks; ideally one would wish to identify shocks experienced by the household head, working age members and also by gender, in line with the existing literature in this area. As in the case of the IHS2 data, income data is also not reported over a period of time which would be preferable, especially when shock data considers a much longer time period. Shocks and income are again self-reported and as such there are potential reliability issues associated with the data.

Future research should consider the timing of shocks as the impact of a shock is likely to dissipate over time. Whilst households may recover from health shocks over a period of a few months or years, in the interim period, the effect on their well-being could be significantly adverse and this needs to be explored further. *Chapter 6* has examined income as an economic outcome in the face of health shocks; other indicators of welfare and well-being should also be considered in future research to examine the impact that health shocks have on livelihoods.

References

Abadie, A and Imbens, G.W. (2006) “Large sample properties of matching estimators for average treatment effects” *Econometrica,* Vol. 74, No.1, pp-235-267.

Angrist, J.D Pischke, JS., (2009) Mostly Harmless Econometrics. An empiricist’s companion. Princeton University Press

Anyamele, O. D. (2011) “The role of wealth in infant mortality in sub-Saharan Africa within urban and between urban and rural areas”, *International Journal of Economics and Research*, Vol.2, No.1, pp.73-81.

Atkinson, A.B and Micklewright, J., (1983) “On the reliability of income data in the family expenditure survey 1970-1977” *Journal of the Royal Statistical Society,* Series A (General), Vol. 146, No. 1, pp.33-61

Austin, P.C., Grootendorst, P. and Anderson, G.M (2006) “A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study”. *Statistics in Medicine,* Vol. 26, No.4. pp.734-753

Barnett, T. and Blaikie, P. (1992) AIDS in Africa: Its present and future impact. Guildford Press, New York.

Barnett, T. and Whiteside, A. (2006) AIDS in the Twenty-First Century. Disease and Globalization, Second Edition, Palgrave Macmillan, Hampshire.

Basu, A.M. and Stephenson, R. (2005) “Low levels of maternal education and the proximate determinants of childhood mortality: A little learning is not a dangerous thing”, *Social Science and Medicine,* Vol.60, No.9, pp.2011-2023.

Becker, G. (1962) “Investment in human capital: A theoretical Analysis” *Journal of Political Economy.* Vol. 70, No.5, p9p9-49.

Becker, G. (1965) “A theory of the allocation of time” *The Economic Journal.* Vol. 75, No. 299, pp.493.517.

Becker, G. (1975) Human capital. A theoretical and empirical analysis with special reference to education. National Bureau of Economic Research. New York. Columbia University Press.

Becker, G., (2007) “Health as human capital: synthesis and extensions” *Oxford University Papers,* Vol.59, pp.379-410.

Beegle K (2003) “Labour effects of adult mortality in Tanzanian households”, *Policy Research Working Paper Series,* No.3062, World Bank.

Ben-Porath, Y. (1967) “The Production of Human Capital and the Life Cycle of Earnings” *The Journal of Political Economy,* Vol. 75. No.4, pp352-365.

Ben-Porath, Y. (1976) “Fertility response to child mortality: Micro data from Israel” *Journal of Political Economy,* Vol. 84, No. 4, pp163-178.

Bernard, H.R, Killworth, P., Kronenfled, D., and Sailer, L. (1984) “The problem of informant accuracy: The validity of retrospective data” *Annual review of Anthropology,* Vol. 13, pp.494-517.

Bicego, G. T. and Boerma, J. T. (1993) “Maternal education and child survival: A comparative study of survey data from 17 countries”, *Social Science and Medicine,* Vol. 36, No.9, pp.1207-1227.

Blane, D., Power, C. and Bartley, M. (1996) “Illness Behaviour and the Measurement of Class Differentials”, *Journal of the Royal Statistical Society, Series A (Statistics in Society),* Vol.159, No.1, pp.77-92.

Blaxter, M. (1989) *“A comparison of measures of inequality in morbidity”* In: Health inequalities in European countries. Edited by Fox, J., Glower Publishing, Aldershot.

Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Gylnn, R.J, Avorn, J and Sturmer, T., (2006) “Variable selection for propensity score models” *American Journal of Epidemiology* Vol.163, No.12, pp1149-1156.

Burbridge, J.B, Magee, L. and Robb, A.L. (1988) “Alternative transformations to handle extreme values of the dependent variable” *Journal of the American Statistical Association,* Vol. 83, No.401, pp.123-127.

Butler, J.S., Burkhauser, R.V. Mitchell, J.M. and Pincus, T..P. (1987). “Measurement Error in Self Reported Health Variables”. *Review of Economics and Statistics*, Vol. 69, No.4, pp644-650.

Caldwell, J. and Caldwell P. (1993) *“Women’s positions and child mortality and morbidity in less developed countries”.* In: Women’s Positions and Demographic Change. Edited by Mason, N. and Sogner, S. Oxford: Clarendon Press.

Caldwell, J. and McDonald, P. (1982) “Influence of maternal education on infant and child mortality: levels and causes”, *Health Policy and Education,* Vol.2, No.3-4, pp. 251-267.

Caldwell, J.C. (1979) “Education as a factor in mortality decline an examination of Nigerian data”, *Population Studies,* Vol.33, No.3, pp.395-413.

Caldwell, J.C. (1994) “How is greater maternal education translated into lower child mortality?", *Health Transition Review,* Vol.4, No.2, pp.224-229.

Caliendo, M and Kopeinig, S (2005) “Some practical guidance for the implementation of propensity score matching” Institute for the Study of Labour. Discussion paper series, IZA DP No.1588

Case, A., Lubotsky, D. and Paxson, C. (2002). “Economic status and health in childhood: The origins of the gradient”. *American Economic Review*, Vol. 92, No.5, pp.1308-1334.

Casterline, J.B., Cooksey, E.C., and Ismail, A.F., (1989) “Household income and child survival in Egypt”, *Demography*, Vol.26, No 1, pp.15-35.

Cleland, J. (2002) “Education and future fertility trends with special reference to mid-transitional countries”. United Nations Population Bulletin. United Nations.

Cleland, J.G. (1990) *“Maternal education and child survival: further evidence and explanations”* In: What We Know about the Health Transition: The Cultural, Social and Behavioural Determinants of Health. Vol.1. Edited by Caldwell, J., Findley, S., Caldwell, P., Santow, G., Braid, J. and Broers-Freeman, D. Canberra Health Transition Centre, The Australian National University.

CSDH (2008) Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health, Geneva, World Health Organization.

Deaton, A. (2003) “Health, Inequality and Economic Development”, *Journal of Economic Literature*, Vol.41, No.1, pp.113-158.

Dehejia, R.H and Wahba, S. (1999) “Propensity Score Matching Methods for Non-Experimental Causal Studies” Disussion Paper No. 0102-14. Columbia University, Department of Economics Discussion Paper Series.

Devereux, S., Baulch, B., Phiri, A. and Sabates-Wheeler, R. (2006) Vulnerability to Chronic Poverty and Malnutrition in Malawi. A Report for DFID Malawi. Institute of Development Studies.

Diamond, P.A. and Hausman, J.A (1984) “Individual retirement and savings behaviour” *Journal of Public Economics*, Vol 23, Issues 1-2, pp. 81-114.

Doepke, M. (2005) “Child mortality and fertility decline: Does the Barro-Becker model fit the facts?” *Journal of Population Economics*, Vol. 18, No. 2, pp337-366.

Donovan, C. Bailey, L. Mpyisi, E and Weber, M. (2003) “Prime-age adult morbidity and mortality in rural Rwanda: Which households are affected and what are their strategies for adjustment?” Proceedings of the 25th International Conference of Agricultural Economists (IAAE), Durban, South Africa.

Dorward A.R. and Mwale, I.M, (2005) *“Labour market and wage impacts of HIV/AIDS in Rural Malawi”* In: AIDS, Poverty and Hunger: Challenges and Responses. International Food Policy Research Institute.

Drummond, M.F., Sculper, M.J., Torrance, G.W. and O'Brien, B.J. (2005) Methods for Economic Evaluation of Health Care Programme. Third Edition, Oxford University Press.

Dufour, J-M, (2003) "Identification, weak instruments and statistical inference in econometrics" *Canadian Journal of Economics,* Vol. 36, Issue 4, pp767-808.

FAO (2004) "Addressing HIV/AIDS through Agriculture and Natural Resource Sectors: a guide for extension workers". Socio Economic and Gender Analysis Programme (SEAGA) Food and Agriculture Organisation of the United Nations. Rome, Italy.

Feeny, D., Furlong, W., Boyle, M. and Torrance, G.W (1995) "Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics*. Vol.7, No.6, pp490.502.

Feng, X. L., Guo, S. Hipgrave, D., Zhu, J. Zhang L., Song, L. Yang, Q. Guo Y and Ronsmans, C., (2011) "China's facility-based birth strategy and neonatal mortality: a population-based epidemiological study", *The Lancet,* Vol.378, No.9801, pp.1493-1500.

Ferro-Luzzi, A. (2003) "Individual food intake survey methods". Keynote paper. Measurement and Assessment of Food Deprivation and Undernutrition. Food and Agriculture Organisation of the United Nations. Rome, Italy.

Gan, L. Xu, L.C and Yao, Y. (2006) "Health shocks, village elections and long-term income: evidence from rural China". NBER Working Paper Series, Working Paper No.12686, National Bureau of Economic Research.

Garcia-Gomez, P., van Kippersluis, H. O'Donnell, O. and Doorslarer, E. (2010) *"Whose Income is least protected against ill health".* In: Understanding Socioeconomic Differences in Health. An Economic Approach. Tinbergen Institute Research Series No.469, Tinbergen Institute.

Ghatak, A. (2010) "Health, labour supply and wages: A critical review of literature". The Institute for Social and Economic Change, Working Paper No.244, Bangalore.

Gold, M.R, Stevenson, D and Fryback, D.G. (2002) "HALYs and QALYs and DALYs, Oh My: Similarities in Summary Measures of Population Health" *Annual Review of Public Health.* Vol 23, pp.115-134.

Grosse, S.D., Lollar, D.J, Campbell, V.A and Chamie, M. (2009) "Disability and disability-adjusted life years: Not the same" *Public Health Reports,* Vol. 124, No.2, pp.197-202

Grossman, M. (1972) "On the concept of health capital and the demand for health" *Journal of Political Economy,* Vol.80, No.2, pp223-255.

Grossman, M. (1999) "The human capital model of the demand for health" *NBER working paper series.* Working paper 7078.

Grosh, M.E. and Grewwe, P. (1995) A guide to Living Standards Measurement Study surveys and their data sets. Policy and Research Department, Poverty and Human Resources Division, Working Paper No.120, World Bank.

Gustafsson-Wright, E., Janssens, W. and van der Gaag, J. (2011) “The inequitable impact of health shocks on the uninsured in Namibia”, *Health Policy and Planning*, Vol. 26, No.2, pp.142-156.

Harrigan, J. (2005) Livelihood Diversification in Malawi. Report of the Scotland Malawi Partnership Conference “Malawi After Gleneagles: a Commission For Africa Case Study”, The Scottish Parliament, Edinburgh, 4-5 November 2005.

Henrich, C., Maffioli, A. and Vazquz, G. (2010) “A primer for applyuing propensity score matching” Impact Evaluation Guidelines, No. IDB-TN-161, Inter-American Development Bank.

Hobcraft, J.N., (1993) “Women's education, child welfare and child survival: a review of the evidence”, *Health Transition Review,* Vol.3, No.2, pp.159-173.

Houweling, T.A.J., Kunst, A.E, Moser, K. and Mackenbach, J.P. (2006) “Rising under-5 mortality in Africa: who bears the brunt?”, *Tropical Medicine and International Health*, Vol.8, No.2, pp.1218-1227.

Houweling, T.A.J and Kunst, A.E (2009) “Socio-economic inequalities in childhood mortality in low- and middle-income countries: a review of the international evidence”, *British Medical Bulletin,* Vol.93, pp7-26.

Huynen, M., Martens, P. and Hilderink, H. (2005) “The health impact of globalisation: a conceptual framework”, *Globalisation and Health*, Vol.1, No.1.

Idler, E.L. and Benyamini, Y. (1997) “Self-rated health and mortality: a review of twenty-seven community studies” Journal of Health and Social Behaviour, Vol. 38, No.1. pp.21-37.

Idler, E.L and Kasl, S.V. (1995) “Self-ratings of health: do they also predict change in functional ability?” *Journals of Gerontology, Series B, Psychological sciences and social sciences.*Vol. 50, No.6, s344-53

Jayne, T.S., Chapoto, A. Byron, E., Ndiyoi, M., Hamazakaza, P., Kadiyala, S and Gillespie, S. (2005) “Community-level Relationships between Prime Age Mortality and Rural Welfare: Panel Survey Evidence from Zambia”. Conference paper presented at the ASSA/AAEA Annual Winter Meetings 6-8 January 2006, Boston MA.

Jeffery, R. and Basu, A (1996) Girls’ Schooling, Women’s Autonomy, and Fertility Change in South Asia. New Delhi, Sage publications.

Johansson, S.R., (1991) “The Health Transition: The Cultural Inflation of Morbidity during the decline of mortality”, *Health Transition Review*, Vol.1, No.1, pp.39-68.

Khandker, S.R., Koolwal, G.B. and Samad, H.A (2010) Handbook on impact evaluation methods. The World Bank. The World Bank, Washington D.C.

KIDS (1998) KwaZulu-Natal Income Dynamics Study (KIDS) Coding Manual. University of Natal, University of Wisconsin and International Food Policy Research Institute.

KIDS (2007) KwaZulu-Natal Income Dynamics Study (KIDS). Third Wave, Public Release Version 1.1. University of Natal, University of Wisconsin and International Food Policy Research Institute.

Kishindo, P. (2004) "Customary land tenure and the new land policy in Malawi", *Journal of Contemporary African Studies.* Vol.22, No.2, pp.213-225.

Kremer, M. and Glennester, R. (2004) Strong Medicine. Creating Incentives for Pharmaceutical Research on Neglected Diseases. Princeton University Press, Oxford.

Lavy, V., Strauss, J., Thomas, D. and de Vreyer, P. (1996) "Quality of health care, survival and health outcomes in Ghana", *Journal of Health Economics*, Vol.15, No.3, pp.333-357.

Lee, L.F (1982) "Health and Wage: A simultaneous equation model with multiple discrete indicators". *International Economic Review,* Vol. 23, No.1, pp. 199-221.

Lee, W-S., (2013) "Propensity score matching and variations on the balancing test" *Empirical Economics*, Vol.44, No.1, pp.47-80.

Lee, S. (2015) "Self-rated health in health surveys". In Johnson, T.P. "Handbook of Health Survey Methods" Wiley Handbooks in Survey Methodology, John Wiley and Sons, Inc, Hoboken, New Jersey.

Loftus, E.F and Marburger, W. (1983) "When personal history repeats itself: Decomposing memories for recurring events" *Applied Cognitive Psychology*, Vol 5, pp. 297-318.

Lucas, H. and Bloom, G. (2006) "Protecting the poor against health shocks". Background paper for "What works for the poorest: knowledge, policies, practices" International conference, BRAC, Rajendrapur 3-5 December 2006.

Macassa, G., Hallqvist, J. and Lynch, J.W. (2011) "Inequalities in child mortality in sub-Saharan Africa: A social epidemiologic framework", *African Journal of Health Science*s, Vol.18, No.1-2, pp14-26.

Manda, S.O.M (1999) "Birth intervals, breastfeeding and determinants of childhood mortality in Malawi", *Social Science and Medicine*, Vol.48, No.3, pp.301-312.

McCollum, E. V. (1939) "Diet and Nutrition Better Nutrition as a health measure" *Canadian Medical Association Journal*. Vol. 40, No.4, pp.393-395.

McDowell, I. (2006) Measuring Health: A guide to rating scales and questionnaires, third edition. Oxford University Press.

Mincer, J.A (1974) Schooling, Experience, and Earnings. NBER Books, National Bureau of Economic Research.

Mosley, W.H and Chen, L.C (1984) “An analytical framework for the study of child survival in developing countries”, *Population and Development Review*, Vol.10, Issue Supplement: Child Survival Strategies for Research, pp.25-45.

Murrugarra, E., and Valdivia, M. (1999) The Returns to Health for Peruvian Urban Adults: Differentials across Genders, the Life-Cycle and the Wage Distribution, Working Paper Series, Working Paper No. 3050, Washington, D.C: Inter-American Development Bank.

Mushkin, S.J., (1962) “Health as an investment” *Journal of Political Economy,* Vol. 70, No.5, pp.129-157.

National Statistics Office Malawi (NSO) (2005) Integrated Household Survey 2004-2005, Vol.1, Household Socio-Economic Characteristics, National Statistics Office Malawi.

Nurse, D. and Philippson, G. (2003) The Bantu Languages. Routledge, New York

ONS (2012) Fertility, 2010-based NPP Reference Volume 3. Office of National Statistics, UK.

Pandey, A. (1998) “Infant and child mortality in India”, National Family Health Survey Subject Reports, No.11, Mumbai, India.

Pebley, A.R and Strupp, P.W (1987) “Reproductive patterns and child mortality in Guatemala”, *Demography,* Vol.24, No.1, pp.43-60.

Pitayanon, S, Kongsin S, and WS Janjareon. (1997) *"The economic impact of HIV/AIDS mortality on households in Thailand,"*. In Bloom, D. and Godwin, P. "The Economics of HIV and AIDS: The Case of South and South East Asia." United Nations Development Programme.

Preston, S.H (2007) “The changing relation between mortality and level of economic development”, *International Journal of Epidemiology*, Vol.36, No.3, pp.484-490.

Psacharopoulos, G and Patrinos, H.A. (2004) “Returns to Investment in Education: A Further Update”, *Education Economics,* Vol.12, No. 2, pp.111-133.

Rahman K.M.M and Sarkar, P (2009) “Levels and differentials of maternal healthcare utilisation in Bangladesh”, *Research Journal of Medical Sciences*, Vol.3, No.4, pp.163-169.

Rajarram, R (2009) Female-Headed Households and Poverty: Evidence from the National Family Health Survey. Federal Research Bank of Atlanta.

Riphahn, R. T. (1999) "Income and employment effects of health shocks. A test case for the German welfare state", *Journal of Population Economics*, Vol.12, No.3, pp.363-389.

Rockwood, T. (2015) "Assessing physical health". In Johnson, T.P. "Handbook of Health Survey Methods" Wiley Handbooks in Survey Methodology, John Wiley and Sons, Inc, Hoboken, New Jersey.

Rosenbaum, P. R. (2002) Observational Studies. Second edition, New York: Springer

Rosenzweig, M.R. and Shultz, P.T. (1983) "Consumer demand and household production: The relationship between fertility and child mortality" *The American Economic Review*, Vol. 73, No. 2 pp38-42.

Sassi, F (2006) "Calculating QALYs, comparing QALY and DALY calculations" *Health Policy and Planning.* Vol.21, No.5, pp.402-408.

Sastry, N. (1997) "What explains rural-urban differentials in child mortality in Brazil?" *Social Science and Medicine*, Vol.44, No.7, pp.989-1002.

Shultz, T. P. (1961) "Investment in human capital" The *American Economic Review*, Vol. 51, No.1, pp.1-17.

Shultz, T.P (2005) "The productive benefits of health. Evidence from low income countries", Economic Growth Center, Yale University, Center Discussion Paper No.903.

Shultz, T.P and Tansel, A. (1997) "Wage and labour supply effects of illness in Côte d'Ivoire and Ghana: instrumental variable estimates for days disabled", *Journal of Development Economics*, Vol.53, No.2, pp.251-286.

Sicotte, M., Ledoux, M., Zunzunegui, MV., Aboubacrine, S.A., Nguyen, VK and the ATARAO group (2010) "Reliability of anthropometric measures in a longitudinal cohort of patients initiating ART in West Africa", *Biomedical Centre Medical Research Methodology*, Vol.10.

Smith, J. A (2000) "A critical survey of empirical methods for evaluating active labour market policies" *Schweizerische Zeitschrift fr Volkswirtschaft und Statistik*, Vol. 136, No.3, pp.1-22.

Smith, J.A and Todd. P.E (2005) "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of Econometrics*, Vol. 125, pp.305-353

Ssewanyana, S. and Younger, S.D. (2005) "Infant Mortality in Uganda: Determinants, Trends and the Millennium Development Goals", *Journal of African Economies*, Vol.17, No.1, pp.34-61.

Staiger, D. and Stock, J.H., (1997) “Instrumental variables regression with weak instruments”, *Econometrica*, Vol.65, No.3, pp.557-586.

Strauss, J. and Thomas, D. (1998). “Health, Nutrition and Economic Development”, *Journal of Economic Literature*, Vol.36, No.2, pp.766-817.

The Roads Authority (2011) Five year strategic and business plan. 2011-2016. Volume 1. The Roads Authority Malawi.

Thomas, D. and Strauss, J. (1997) “Health and wages: evidence on men and women in urban Brazil”, *Journal of Econometrics*, Vol.77, No.1, pp.159-185.

Thompson, F.E., Moler, J.E., Freedman, L.S., Clifford, C.K., Stables, G.J. & Willet, W.C. 1997. “Register of dietary assessment calibration-validation studies: a status report”, *American Journal of Clinical Nutrition*, Vol.64, No.4 (supplement), pp.1142 - 1147.

Torrance, G.W. and Feeny, D. (1989) “Utilities and quality adjusted life years” *International Journal of Technology Assessment in Health Care.* Vol. 5, No.4, pp.559-75

United Nations (2004) The Impact of AIDS, Department of Economic and Social Affairs, Population Division, United Nations.

UNICEF (2012) Levels and Trends in Child Mortality, Report 2012. Estimates developed by the UN Inter-agency Group for Child Mortality Estimation. United Nations Children’s Fund, New York.

Urassa, M., Boerma, J.T., Isingo, R., Ngalula, J., Ng'weshemi, J., Mwaluko, G., and Zaba, B. (2001) “The impact of HIV/AIDS on mortality and household mobility in rural Tanzania”, *AIDS*, Vol.15, No.15, pp.2017-2023.

UNAIDS (2005) The Governance Dimensions of Food Security in Malawi, US Agency for International Development.

UNAIDS (2010) User Guide: CSPRO data entry and analysis software. For implementers of USAID poverty assessment tools, July 2010. US Agency for International Development, IRIS Center, University of Maryland.

UNAIDS (2011) World AIDS day report 2011. “How to get to zero: Faster. Smarter. Better.” Joint United Nations Programme on HIV/AIDS. US Agency for International Development.

UNAIDS (2012) Global Report. UNAIDS report on the Global AIDS Epidemic 2012. Joint United Nations Programme on HIV/AIDS (UNAIDS).

Van Doorslaer, E and Jones, A.M (2003) “Inequalities in self-reported health: validation of a new approach to measurement”, *Journal of Health Economics*, Vol 22, pp.61-87.

Wagstaff, A. (2007) “The economic consequences of health shocks: Evidence from Vietnam”, *Journal of Health Economics*, Vol.26, No.1, pp.82-100.

Wang, L. (2003) “Determinants of child mortality in LDCs: Empirical findings from demographic and health surveys”, *Health Policy*, Vol.65, No.3, pp.277-299.

Weinstein, M.C., Torrance, G.T and McGuire, A., (2009) “QALY: The Basics” *Value in Health.* Vol 12. Supplement 1, s.5-9.

Webb, G. P. (2007) Nutrition A Health Promotion Approach, Third Edition, Taylor Francis Group.

Whiteside, M. (2000) Ganyu labour in Malawi and its implications for livelihood security interventions. An analysis of recent literature and implications for poverty alleviation. AgREN Network Paper, No.99, Overseas Development Institute.

WHO (2008) Global Health Observatory Data Report, World Health Organization.

WHO (2011) World Malaria Report 2012, WHO Global Malaria Programme, World Health Organization.

Willis, R.J (1986) “Wage determinants: a survey and reinterpretation of human capital earnings functions” In: Handbook of Labour Economics, Vol.1 edited by Ashenfelter, O. and Layard, R. Elsevier Science Publishers.

Wooldridge, J.M. (1999) Introductory Econometrics: A modern approach. South-Western College Publishing.

World Bank (1993) World Development Report, 1993. Investing in Health, World Bank, New York, Oxford University Press.

World Bank (2006) Poverty and vulnerability assessment. Investing in our future, Volume 1, Synthesis of Main Findings and Recommendations. World Bank Report No. 36546, Poverty Reduction and Economic Management 1, Africa Region.

World Bank (2012) World Bank Development Indicators, World Bank, New York.

Yamano, T. and Jayne T.S (2004) “Measuring the Impacts of Working-Age Adult Mortality on Small-Scale Farm Households in Kenya”, *World Development*, Vol.32, No.1, pp.91-119.

Zeckhauser, R and Shepard, D. (1976), “Where Now for Saving Lives?” *Law and Contemporary Problems*, Vol. 40, No. 4 pp.5-45